

UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR



**Grado en Ingeniería de Tecnologías y Servicios de
Telecomunicación**

TRABAJO FIN DE GRADO

**DETECCIÓN Y SEGUIMIENTO DE VEHÍCULOS
MEDIANTE ANÁLISIS DE IMÁGENES**

Luis Miguel Gil Ramos

Tutor: Luis Salgado Álvarez de Sotomayor

Ponente: Jesús Bescós Cano

JULIO 2016

Resumen

El objetivo de este Trabajo de Fin de Grado es la detección de la posición de un vehículo en relación al centro de su carril. Dicho vehículo obtiene información de su entorno por medio de una cámara situada en la parte frontal.

Para ello, se plantea una estrategia basada en la segmentación de la imagen obtenida en tres clases: línea, pavimento y objeto. Se estudiarán distintos métodos para clasificar los píxeles de la imagen de acuerdo a esas tres clases, todos ellos basados en el conocimiento estadístico de algunas de las características propias de imágenes de carretera.

Uno de los métodos que utilizaremos se basa en estadística bayesiana, que nos permitirá añadir conocimiento que conocemos (llamado *a priori*) para mejorar la segmentación. Usaremos esta posibilidad para mejorar el sistema a través de la introducción de conocimiento *a priori* basándonos en el movimiento.

Todas estas etapas se han estudiado finalmente de acuerdo a unas pruebas presentadas al final del capítulo, en las que se compararán los métodos de segmentación con imágenes etiquetadas para este propósito. Se analizarán las mejoras propuestas de acuerdo al fin de este Trabajo de Fin de Grado como para otro tipo de aplicaciones.

Palabras clave:

Modelado del entorno, detección del pavimento, líneas de carretera y vehículos, Transformación Inversa de Perspectiva, clasificación bayesiana, algoritmo Expectation-Maximization, segmentación, algoritmo KLT.

Abstract

The aim of this work is detecting the position of a vehicle relative to the center of your lane. This vehicle obtains information from its environment by a camera located on the front.

For that, we use a strategy based on image segmentation in three classes which are line, pavement and object. Different methods to classify image pixels according to these three classes will be studied, all of them based on statistical knowledge of some of the characteristics of road images.

One method we use is based on Bayesian statistics, which will allow us to add knowledge we know (called *a priori*) to improve segmentation. We will use this to improve the system through the introduction of *a priori* knowledge based on the movement.

All these steps have finally studied according to some evidence presented at the end of the work, in which segmentation methods will be compared with images labeled for this purpose. the proposed improvements according to this Work to End of Grade and for other applications will be discussed.

Keywords:

Environment modeling, detection pavement, road and vehicle lines, Inverse Perspective Transformation, Bayesian classification, Expectation-Maximization algorithm, segmentation, KLT algorithm.

Agradecimientos

A mi familia, siempre dispuesta, siempre conmigo, siempre encima.

Al año y medio difícil que he pasado, por poder tratarlo como tal.

A las amistades que están, los amores que van y vienen, las noches perdidas por las calles de Madrid.

A una vida que sigue, que merece la pena en las malas, aún más en las buenas.

A los profesores que, desde que fui un crío, fueron enseñándome más allá de números y letras, y en un lugar mucho más cercano como es nuestro corazón, y humanidad.

Y los primeros de todos, mis queridos abuelos.

Incluso a un tutor que me ha aguantado como nadie, en mis buenas y en todas mis malas, llamado Luis Salgado.

A los demás profesores, en su justa medida, pero,

A todos, personas, momentos e ideas,

Hasta a un pequeño pueblo como Barruelo de Santullán,

Luis Miguel Gil Ramos

Julio, 2016

Índice general

Palabras clave	III
Resumen	III
Keywords	V
Abstract	V
Agradecimientos	VII
Índice general.....	IX
Índice de tablas	XII
Capítulo 1: Introducción	1
1.1 Motivación.....	1
1.2 Objetivos	2
1.3 Organización de la memoria	2
Capítulo 2: Estado del Arte	3
2.1 Introducción	3
2.2 Detección de líneas de carril	5
2.3 Detección de vehículos.....	6
2.4 Transformada de Perspectiva Inversa (IPM)	7
Capítulo 3: Arquitectura del sistema	9
3.1 Introducción	9
3.2 Modelado de la imagen.....	11
3.2.1 Modelado basado en filtro Detector de Líneas de Carretera (Lane Marking Detector, LMD)	12
3.2.2 Modelado basado en Características de Intensidad (Intensity Features Extractor, IFE)	15
3.3 Refinamiento del modelado mediante el Algoritmo Expectation- Maximisation (EM).....	19
3.4 Clasificación de la imagen utilizando Máxima Verosimilitud o Máxima a Posteriori	19
3.4.1 Clasificación basada en Máxima Verosimilitud.....	20
3.4.2 Clasificación basada en Máxima a posteriori.....	20
3.4.2.1 Generación de <i>prioris</i>	21
3.5 Control de segmentación	24
3.6 Detección de posición.....	26
Capítulo 4: Pruebas y resultados	27
4.1 Introducción	27
4.2 Medidas.....	28

4.2.1 Clase Pavimento	28
4.2.2 Clase Objeto	29
4.2.3 Clase Línea	30
4.3 Resultados	31
4.3.1 Clase Pavimento	31
4.3.2 Clase Objeto	33
4.3.3 Clase Línea	37
4.3.4 Análisis subjetivo	39
Capítulo 5: Conclusiones y trabajo futuro	41
Referencias	43
Anexos	47
Anexo 1. Estadística Bayesiana y Funciones de verosimilitud	47
Anexo 1.1 Funciones de verosimilitud	48
Anexo 1.2 Decisión bayesiana	49
Anexo 2. Kanade-Lucas-Tomasi (KLT)	50

Índice de tablas

Figura 3.1 Diagrama de bloques del sistema	10
Figura 3.10 Obtención de las máscaras gaussianas.....	24
Figura 3.11 Posición del centro del carril respecto al vehículo.....	26
Figura 3.2 Imagen original y su visualización posterior a la aplicación del filtro de líneas	13
Figura 3.3 Histograma de la respuesta del filtro de línea. Se corresponde a la imagen de la figura 3.2.b.....	13
Figura 3.4 Modelo obtenido a partir de los parámetros del LMD, que caracterizarían las dos clases a definir. En rojo, la clase oscura (objeto y pavimento); en azul, la clase línea	14
Figura 3.6 Obtención de la máscara del vehículo por Otsu y combinación de los dos métodos	17
Figura 3.7 Gaussianas obtenidas a partir de los parámetros del IFE, que caracterizarían las tres clases a definir antes de normalizarse. En rojo, la clase objeto; en verde, la clase pavimento; en azul, la clase línea.	18
Figura 3.8 Obtención de puntos con los que inicializaremos el KLT.....	22
Figura 3.9 Obtención de las gaussianas, previo paso a la obtención de máscaras	23
Figura 4.1 Imagen original y su ground truth.....	28
Figura 4.10 Diferencias en píxeles entre el comienzo de los vehículos detectados y los de referencia para la región de interés	37
Figura 4.11 Segmentos de línea detectados entre los cinco sistemas y la imagen de referencia	38
Figura 4.12 Ángulos que parametrizan la orientación de las líneas de carretera.....	38
Figura 4.13 Ejemplo de una imagen IPM segmentada con diversos métodos propuestos	39
Figura 4.14 Ejemplos de la secuencia de referencia utilizada para la detección del cambio de carril	40
Figura 4.15 Posición del centro del carril respecto al vehículo.....	40
Figura 4.2 Ejemplo de la geometría de un vehículo debido a la transformada IPM.....	30
Figura 4.3 Puntuaciones F-Score clase pavimento.....	31
Figura 4.4 Imagen original e imagen segmentada para el peor resultado F- Score obtenido.....	32
Figura 4.5 Ángulos que parametrizan los laterales de los vehículos	33

Figura 4.6 Ejemplo de una imagen de referencia a partir de la que comparamos las imágenes segmentadas para cada uno de los métodos. Nótese la semejanza en la geometría del vehículo en todos los casos	34
Figura 4.7 Imágenes obtenidas de la secuencia de referencia por orden cronológico.....	35
Figura 4.8 Diferencias en el ángulo obtenido para las rectas que caracterizan los laterales, ampliado en la zona que produce el cambio de orientación	36
Figura 4.9 Diferencias en píxeles entre el comienzo de los vehículos detectados y los de referencia para toda la imagen	36
Figura Anexo 2.1 Esquema del funcionamiento de la estadística	47

Capítulo 1: Introducción

1.1 Motivación

En el contexto de sistemas de ayuda a la conducción¹, determinar la posición de nuestro vehículo en la carretera, así como la del resto de vehículos en nuestro entorno, es uno de los objetivos más importantes [1]. Los cambios de condiciones, como pueden ser los cambios de iluminación, de la geometría de las líneas de la carretera, de la textura del pavimento, etcétera, condicionarán este tipo de aplicaciones.

Además, nuevas formas de relacionarse entre vehículos, conductores y la propia infraestructura son posibles gracias a los sistemas inteligentes de transporte², los cuáles necesitarán información precisa de las condiciones de la carretera, los vehículos que hay en cada momento en ella y su lugar, etcétera.

Todo ello en realidad se preocupa por un mismo objetivo: la seguridad basada en la convivencia de todos los aspectos implicados en la conducción. De acuerdo a [2], el número de muertes en carretera en 2015 es de 1.25 millones, lo que nos lleva a buscar formas con las que reducir esta tasa.

Algoritmos y métodos bien conocidos se han ido desarrollando de una forma exponencial para hacer que los sistemas de ayuda a la conducción y los sistemas inteligentes de transporte sean cada vez más robustos. El análisis de vídeo basado en sensores pasivos (es decir, aquellos que no emiten señales ni interfieren con el medio, que sería el caso de los sensores activos), ha tomado un papel fundamental en este tipo de sistemas, si bien un amplio número de soluciones basadas en la combinación de ambos tipos de sensores han sido discutidas en diversos artículos.

¹ Advanced Driver Assistance Systems, ADAS.

² Intelligent Transportation System, ITS.

1.2 Objetivos

El objetivo de este Trabajo de Fin de Grado es el cálculo de la posición del vehículo en la carretera, siendo capaces de detectar un cambio de carril y en qué dirección se está dando.

Para ello, haremos uso de la implementación y análisis de un sistema capaz de segmentar las imágenes de carretera obtenidas por una cámara situada en la parte frontal del vehículo, mirando a la carretera.

Este sistema basado en el propuesto en [3] tratará de caracterizar el entorno del vehículo basándonos en imágenes conocidas como IPM³, que transforman la visión espacial de la cámara en una visión cenital.

Posteriormente, el sistema utilizará un modelo de clasificación y segmentación estadística de la imagen, basado en tres clases, que definiremos como: objeto, pavimento y línea.

Por último, se analizarán los resultados de los diferentes métodos propuestos para la segmentación, así como de la detección del cambio de carril. Nos basaremos en dos tipos de pruebas, objetivas de acuerdo a las propuestas en [4] y subjetivas.

1.3 Organización de la memoria

La memoria presenta la siguiente estructura:

- ❖ Capítulo 1: Introducción
- ❖ Capítulo 2: Estado del Arte
- ❖ Capítulo 3: Arquitectura del sistema implementado
- ❖ Capítulo 4: Pruebas y análisis de resultados
- ❖ Capítulo 5: Conclusiones y trabajo futuro

³ Por provenir de la Transformada Inversa de Perspectiva (*Inverse Perspective Mapping, IPM*). Se explica en el apartado 2.4 dentro del Estado del Arte

Capítulo 2: Estado del Arte

2.1 Introducción

Este Trabajo de Fin de Grado se ha podido desarrollar en gran medida gracias al aumento de la potencia de los microprocesadores, debido a la alta velocidad de procesamiento que requieren las aplicaciones de análisis de vídeo [5]. Sumado a sensores capaces de capturar vídeo digitalmente, la posibilidad de extraer y analizar información del entorno de nuestro vehículo se convierte en un hecho factible.

Hay varias formas de enfocar el problema que supone conseguir datos sobre la situación de nuestro vehículo en la carretera, así como de dónde se encuentra en relación a otros vehículos y el comportamiento de éstos.

Una de esas formas incluye el uso de sensores pasivos, como las cámaras, que no son intrusivas con el medio. Trabajos como [6] utilizan sistemas mono-cámara para el posterior análisis de las imágenes en tiempo real.

Por otro lado, están los sensores activos, como sónares y radares, que interactúan de alguna forma con el entorno, por lo general mediante emisión de señales y recepción de las mismas después de haberse transformado por el medio. Un ejemplo es [7], en el que una serie de sensores infrarrojos situados en la parte frontal de nuestro vehículo sería capaz de detectar peatones.

A su vez, dentro de los trabajos basados en la utilización de cámaras, podemos encontrar dos grandes grupos:

- ❖ Trabajos centrados en cámara fija, en los que, como su nombre indica, no hay movimiento del dispositivo. Las imágenes del entorno obtenidas suelen usarse en distintos campos: detección de accidentes de carretera, incumplimiento de leyes, etcétera. En [8] se propone el uso de una cámara fija para el conteo de vehículos.
- ❖ Cámara introducida en el vehículo. Detecta las reacciones y estímulos del entorno en el que es montada. Por lo general, estará en la parte frontal o en el techo del vehículo. Varias aproximaciones, que se pueden agrupar:

- Según el número de cámaras. En [9] o en [10] se proponen sistemas estéreo, en los que las distancias entre objetos y la profundidad de la escena se pueden extraer con mayor facilidad. Por otro lado, los sistemas mono-cámara son más sencillos y no requieren de calibrado, como los propuestos en [3] , [8] o [6].
- Según el número de regiones a detectar. Pueden etiquetar los diferentes elementos de una escena, como peatones, entorno del vehículo, etcétera. Un trabajo de este tipo se recoge en [11], en el que se etiqueta de forma semántica la escena incluyendo peatones, líneas de carreteras y vehículos, entre otros.
- Según el elemento único que se desea detectar. Trabajan de forma precisa a la hora de extraer la información para un solo elemento de la escena. Por ejemplo, en [12] se detectan exclusivamente las líneas de carretera, mientras que otros trabajos como [13] se centran en detectar únicamente los vehículos.

El objetivo de este Trabajo de Fin de Grado requiere extraer información del entorno previamente para poder caracterizar la carretera y detectar vehículos, así como determinar el comportamiento de nuestro vehículo.

Para ello, utilizaremos una única cámara montada en nuestro vehículo. La elección de este tipo de dispositivos conlleva dos grandes problemas: la detección de los elementos estáticos de la carretera, como pueden ser las líneas, pavimento y los objetos no dinámicos, y la detección de vehículos. Ambos han sido tratados con anterioridad en otros trabajos, como veremos a continuación.

Sin embargo, las imágenes que obtengamos por la cámara no las usaremos directamente para analizar la información. Antes, realizaremos una Transformada Perspectiva Inversa (*Inverse Perspective Mapping, IPM*) a las imágenes obtenidas por la cámara, ya que nos proporcionará una vista cenital en la que extraer los datos necesarios es más fácil. Por ello, también hablaremos de la transformación IPM en este capítulo.

2.2 Detección de líneas de carril

Punto fundamental para los métodos de ayuda a la conducción. Con ello, podemos analizar la evolución de la conducción y hacer que el vehículo sepa por dónde circular.

Se define la detección de líneas de carril como el problema para localizar los límites del carril sin conocimiento previo de la geometría de la carretera [14]. Después, se puede generar un modelo que las describa para poder hacer un seguimiento imagen a imagen.

Se pueden encontrar dos tipos de aproximaciones diferentes a este tipo de detección:

- ❖ Basadas en patrones: más robusto frente al entorno, no dependen del estado de las líneas de la carretera o las condiciones de iluminación. Sin embargo, son menos flexibles ante nuevas formas que no estén en el diseño. En [15] se extraen características de la imagen y se comparan con patrones de forma estadística, para generar un modelo matemático que las describa.
- ❖ Basadas en características, las cuáles a su vez se pueden dividir en tres grandes grupos:
 - Usando detectores de bordes, muy afectado por cambios de iluminación. Además, es difícil encontrar un umbral de intensidad capaz de excluir aquellas componentes no pertenecientes a línea, como se presenta en [16].
 - A través del estudio y aprovechamiento de los histogramas. Haciendo asunciones sobre la forma esperada del histograma, concretamente que los píxeles con mayor intensidad pertenecerán a las líneas, se obtienen aquellas regiones probables de ser línea. Un ejemplo de este tipo de trabajos se muestra en [6].
 - Asumiendo que las rectas tienen un ancho constante gracias al uso de transformadas como la IPM. A través de la aplicación de filtros morfológicos, se pueden obtener altas respuestas a las líneas, como se presenta en [10].

Debido a la simplicidad y buen rendimiento que obtiene, en este TFG se usará una detección basada en características. Particularmente, se trabajará sobre el dominio transformado IPM, por lo que podremos asumir que las líneas tienen un ancho razonablemente constante.

2.3 Detección de vehículos

Generalmente constan de tres etapas: generación de hipótesis, verificación de hipótesis y seguimiento del vehículo.

En la generación de hipótesis, se recorre la imagen para obtener potenciales regiones de contener vehículos. La forma de analizar la información de la imagen para generar las hipótesis puede dividirse en:

- ❖ Visión estéreo: explotan la capacidad para detectar profundidad gracias a la diferencia entre imágenes captadas por cada cámara. La información métrica de los objetos se extrae fácilmente. Trabajos como [9] utilizan este tipo de hipótesis para buscar vehículos en la imagen.
- ❖ Basadas en movimiento: se analizan los vectores de movimiento entre dos imágenes consecutivas. Se utilizan y explican en [17], remarcando el enorme coste computacional que requiere. En dicho trabajo, lo solucionan calculando los vectores de movimiento únicamente en los bordes.
- ❖ Basadas en conocimiento: hacen uso de información previa para localizar vehículos en la imagen. Por ejemplo, la simetría de los vehículos en su parte trasera, sombras que definen regiones oscuras, etcétera. Técnicas basadas en simetría y brillos se utilizan en [9].

En la segunda etapa, de verificación de las hipótesis, el objetivo será validar o rechazar los candidatos propuestos en la generación de hipótesis. De nuevo, se pueden ver varias formas de realizar este proceso:

- ❖ Basadas en modelos: utilizan modelos de los vehículos para poder realizar medidas de similitud. Bordes o simetrías suelen utilizarse en este tipo de verificación, como ocurre en [18].
- ❖ Basadas en aprendizaje: se basa en la clasificación de las hipótesis a partir de un conjunto de ejemplos de entrenamiento. Redes neuronales (*Neural Networks, NN*) o Máquinas de Vectores de Soporte (*Support Vector Machines, SVMs*), se utilizan para este tipo de verificaciones, como se explica en [19].

En este trabajo, se utilizarán técnicas basadas en movimiento y en conocimiento para la generación de hipótesis. Concretamente, una característica robusta que se mantiene en muchas situaciones es la sombra inferior del vehículo, que puede usarse para su detección. Este concepto fue introducido en [20] y utilizado en otros trabajos con un buen resultado, como en [16] o en [21].

La etapa de verificación de hipótesis se introduce en la etapa de pruebas, basándonos en modelos. Concretamente, disponemos de unas imágenes IPM

etiquetadas píxel a píxel según su clase a mano (*ground truths*), que usaremos para estudiar el rendimiento de nuestro sistema.

Puesto que en este Trabajo de Fin de Grado no se realiza un seguimiento del vehículo, sólo destacar que hay un amplio uso en el Estado del Arte de seguimiento basado en aproximaciones bayesianas. Dentro de este tipo de aproximaciones, destacar las que hacen uso del filtro Kalman, como [3], que asumen modelos lineales de la velocidad y la aceleración para predecir la situación del vehículo.

2.4 Transformada de Perspectiva Inversa (IPM)

Introducida por Bertozzi y Broggi en [10], es una transformada geométrica diseñada para eliminar los efectos de perspectiva en las imágenes adquiridas por una cámara situada en el frontal del vehículo. De esta forma, obtenemos una visión cenital de la escena, tal y como podemos observar en la figura 2.1:



a) Imagen original

b) Imagen IPM

Figura 2.1 Obtención de una imagen IPM

Esta transformada presenta una serie de ventajas que la hacen de especial utilidad en los sistemas de detección basados en visión. Gracias a que no se muestrea de forma homogénea la imagen de entrada para generar la transformada [22], las líneas de carretera presentan un ancho constante además de ser paralelas, propiedades muy útiles en la detección y modelado de líneas de carril. A su vez elimina uno de los efectos de la perspectiva, por el cual el mismo número de píxeles en la imagen original no representa el mismo espacio en el dominio real, haciendo que el cálculo de la velocidad y posición del vehículo sea más sencillo. Por contrapartida, cualquier elemento que tenga una altura mayor a la de la carretera se verá deformado, por lo que los vehículos pierden su forma y propiedades de simetría en este dominio.

Originalmente esta transformación requería saber las condiciones específicas de adquisición: posición de la cámara, orientación, etc. además de asumir que

las carreteras eran totalmente planas. Posteriores trabajos relajaron esos requisitos, automatizando parte del proceso de calibración a partir del análisis de las imágenes de entrada. Así por ejemplo en [23] se elimina la distorsión y se proponen formas de obtener los parámetros necesarios de una forma controlada.

Capítulo 3: Arquitectura del sistema

3.1 Introducción

El objetivo del sistema es la detección de los cambios de carril de nuestro vehículo, así como la posición de éste dentro del carril.

Para ello, se desarrollarán técnicas para clasificar los píxeles de las imágenes tomadas por la cámara colocada en el frontal de nuestro vehículo según pertenezcan a las clases de interés: líneas de carretera, pavimento y objeto. Éstas se pueden definir como:

- ❖ Objetos: elementos de interés (vehículos) caracterizados fundamentalmente por la sombra que proyectan en el pavimento.
- ❖ Pavimento: zonas grises de bajo gradiente. Por lo general, será la clase predominante de la imagen.
- ❖ Línea: zonas de alta intensidad lumínica conformada por líneas o rectángulos. Presentarán un alto gradiente con respecto a su entorno, que será de clase pavimento u objeto. Esta información es la más susceptible de perderse, la más escasa y la que más nos interesa para nuestros objetivos.

De esta manera, obtendremos información de la ubicación de las líneas de carretera (clase línea) que utilizaremos para nuestro objetivo.

Adicionalmente, tendremos información del entorno de la carretera, como regiones potenciales de contener vehículos (clase objeto).

En la figura 3.1 se presenta un diagrama de bloques del sistema desarrollado:

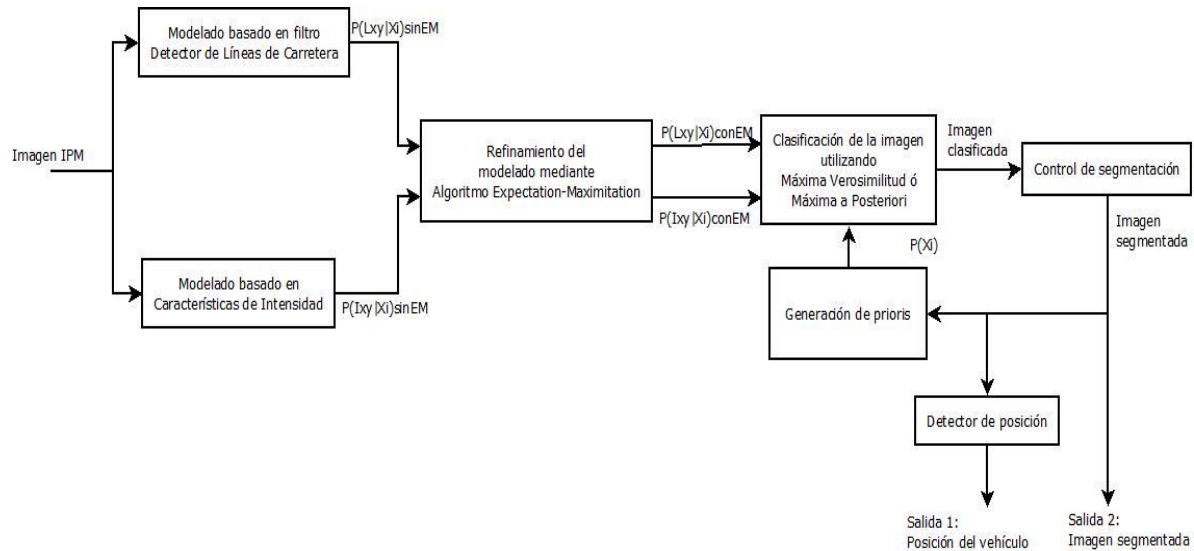


Figura 3.1 Diagrama de bloques del sistema

De acuerdo con dicho el diagrama, la entrada del sistema son las imágenes IPM ya presentadas. Con ellas, se modelarán las clases objeto, pavimento y línea de manera paramétrica, empleando para ello un filtro Detector de Líneas de Carretera (*Lane Marking Detector, LMD*) y empleando el nivel de intensidad asociado a cada una de las clases a través de un extractor de parámetros basado en Características de Intensidad (*Intensity Features Extractor, IFE*).

La salida de ambos módulos de procesamiento será la probabilidad de cada píxel de pertenecer a una clase u otra, pero de acuerdo a las clases que propone cada modelo. Con esas probabilidades, se pueden tomar hasta cuatro decisiones:

1. Clasificación de la imagen según el principio de Máxima Verosimilitud con el modelo sin refinar: las probabilidades para cada clase se toman como el producto de las probabilidades independientes entre sí del LMD y del IFE. La clase de cada píxel será aquella cuya mayor probabilidad presente.
2. Clasificación de la imagen según el principio de Máximo a Posteriori con el modelo sin refinar: al igual que en el caso anterior, las probabilidades para cada clase se toman como el producto de las probabilidades independientes entre sí del LMD y del IFE. Sin embargo, la clase de cada píxel no será la mayor de las probabilidades, sino que responderá además a un conocimiento *a priori* introducido por medio de la probabilidad bayesiana.
3. Clasificación de la imagen según el principio de Máxima Verosimilitud con modelos refinadas vía el algoritmo EM: las medias y varianzas

obtenidas en el LMD y el IFE se reajustan por medio del algoritmo *Expectation-Maximitation(EM)*. Después, se haría lo mismo que en el caso 1.

4. Clasificación de la imagen según el principio de Máximo a Posteriori con modelos refinados vía el algoritmo EM: una combinación de los casos 2 y 3. Las medias y varianzas ajustadas a través del algoritmo EM se utilizan para clasificar la imagen de acuerdo con la teoría de Bayes.

Todas las salidas anteriores serán discutidas entre sí, comparando su rendimiento general y por clases en el apartado de pruebas.

En la fase de control de segmentación se utilizarán una serie de mecanismos para verificar que los parámetros son correctos y dar robustez al sistema, introduciéndose también conocimiento adicional para mejorar la segmentación.

Por último, se medirá la posición de nuestro vehículo en la carretera, pudiendo detectar cambios de un carril a otro. Para ello, la segmentación debe de ser robusta y fiable al menos en una región de interés. Sabiendo que la transformada IPM añade mucho ruido en la zona alta y los laterales de la imagen, nos quedaremos con la parte central-inferior de la imagen.

Los mayores avances de este TFG se verán reflejados sobre todo en la parte de generación de información *a priori* para la decisión bayesiana, basándonos en el movimiento del vehículo. La introducción de este tipo de información también será discutida en el apartado de pruebas.

3.2 Modelado de la imagen

La manera en la que vamos a decidir la pertenencia de un píxel a una clase u otra estará basada en métodos estadísticos. Este tipo de elecciones basadas en probabilidad describen la información del entorno de una manera simple y robusta [3], lo que los hace muy prácticos.

Sin embargo, requieren que se escoja de una manera adecuada las características y muestras que representen a la población.

El objetivo de este bloque es generar un modelo fiable y sencillo que nos permita una clasificación lo más fiel posible. Para este propósito, se utilizará un modelado gaussiano de las clases propuestas en cada modelo, por dos razones: los modelos gaussianos han ofrecido buenos resultados en estado del

arte, como en [21], así como su sencillez, dado que permiten describir una clase con sólo dos parámetros, la media y la varianza.

De forma general, una función de densidad gaussiana, más conocidas como distribuciones normales, responde a la siguiente fórmula (caso unidimensional):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (1)$$

donde σ es la desviación típica de las muestras x y μ es su media.

No habrá por ello que olvidar que, aunque hablemos de medias y varianzas, en realidad estamos trabajando con un modelo gaussiano, asignando a cada píxel unas probabilidades de pertenencia a cada clase.

3.2.1 Modelado basado en filtro Detector de Líneas de Carretera (*Lane Marking Detector, LMD*)

Es uno de los dos extractores de parámetros en función de un modelo gaussiano. Se basa en el conocimiento de las líneas como zonas estrechas y de una intensidad alta respecto a su entorno (pavimento u objeto), más oscuro. Por ello, permite modelar la probabilidad de que un píxel pertenezca a cada clase en función de su respuesta a un filtro cuya ecuación se define como:

$$y_i = 2x_i - (x_{i-\tau} + x_{i+\tau}) - |x_{i-\tau} - x_{i+\tau}|$$

donde y_i es el píxel de la fila i en la imagen filtrada, x_i el píxel de la fila i en la imagen de entrada IPM, y τ es el ancho estimado de la línea en píxeles.

La idea del filtro es resaltar en el primer término el píxel de la fila i y restarlo con sus vecinos de fila τ columnas a la izquierda y τ columnas a la derecha. De esta manera, píxeles con alta intensidad con un entorno a ambos lados oscuro (lo que ocurre en el caso de líneas de carretera) dará una respuesta alta, mientras que píxeles en una región homogénea, como es el caso del pavimento, darán una respuesta baja.

Un ejemplo de aplicación de este filtro es el mostrado en la figura 3.2:



a) Cuadro IPM Original



b) Imagen resultado de aplicar el filtro de líneas

Figura 3.2 Imagen original y su visualización posterior a la aplicación del filtro de líneas

Como puede verse en la imagen, el LMD es muy fiable para la detección de líneas de carretera. También podemos observar que presenta respuestas altas en zonas ruidosas de la imagen, como la parte superior y los laterales. Al ser ruido proveniente de la transformación IPM, asumiremos información fiable aquella en la zona central-inferior de la imagen, que se correspondería a la más cercana a nuestro vehículo.

Un ejemplo del número de píxeles que presentan una u otra respuesta puede verse en el histograma de la figura 3.3:

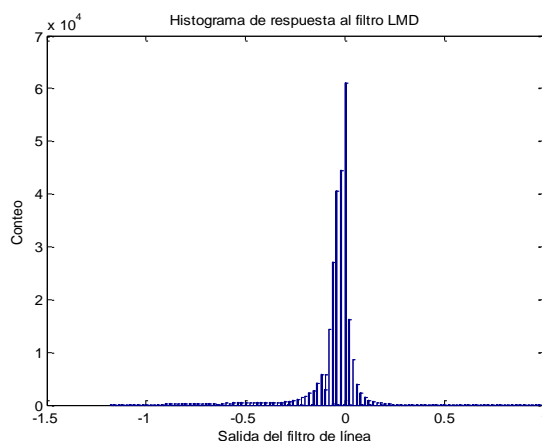


Figura 3.3 Histograma de la respuesta del filtro de línea. Se corresponde a la imagen de la figura 3.2.b

Observando la figura 3.3, vemos cómo un gran número de píxeles tienen una respuesta baja entorno al cero, lo que concuerda con la idea de que la mayoría de los píxeles se encuentran en regiones homogéneas. Píxeles con respuestas muy bajas indican que se encuentran en regiones oscuras y estrechas rodeadas de regiones de intensidad alta, mientras que los píxeles con

respuestas mayores que cero serán píxeles de una intensidad mayor que la de su entorno, como ocurre con las líneas.

Utilizaremos un modelo de dos clases para representar las muestras: una clase con respuesta alta cuya población serán píxeles de línea, y una clase con respuesta baja para píxeles de objeto y pavimento, sin poder distinguirlos entre sí ya que la respuesta del filtro en ambos casos es muy similar.

Para ello, empleamos un umbral situado en la media de la respuesta al filtro más la desviación típica, dado a que ofrece un buen rendimiento a la hora de separar la población de cada clase, y se ajusta a lo deducido a partir del histograma. Dicho umbral ha sido probado empíricamente dentro de unos rangos lógicos⁴.

Los parámetros que extraeremos, por tanto, serán las medias y las varianzas de las muestras (píxeles) que queden a los lados del umbral fijado.

Sustituyendo dichas medias y varianzas en la ecuación (1), habríamos modelado las dos clases. La notación utilizada para describir la probabilidad de un píxel de pertenecer a una u otra clase será la siguiente:

- ❖ $P(L_{xy}|X_{hom})$, o probabilidad de que un píxel tenga una respuesta L_{xy} al filtro sabiendo que pertenece a la clase de respuesta baja u homogénea, X_{hom} .
- ❖ $P(L_{xy}|X_{lin})$ o probabilidad de que un píxel tenga una respuesta L_{xy} al filtro sabiendo que pertenece a la clase de respuesta alta siendo muy probablemente línea, X_{lin} .

Las probabilidades $P(L_{xy}|X_i)$ se conocen como funciones de verosimilitud, y vienen explicadas en el Anexo 1.

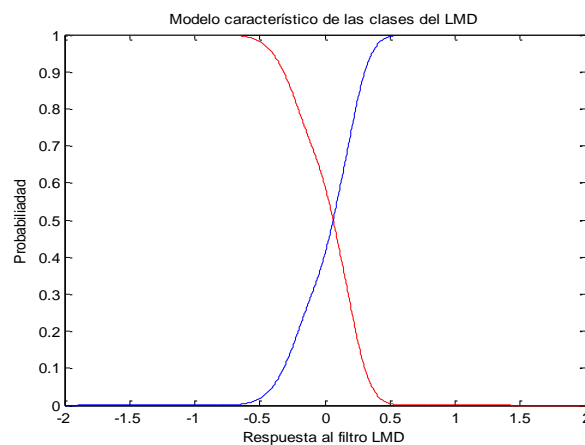


Figura 3.4 Modelo obtenido a partir de los parámetros del LMD, que caracterizarían las dos clases a definir. En rojo, la clase oscura (objeto y pavimento); en azul, la clase línea

⁴ Por encima de la media de la respuesta y por debajo de la respuesta máxima.

Para concluir, se presenta en la figura 3.4 el resultado de este tipo de modelado, una vez que se normalizan las probabilidades de las gaussianas para cada píxel. Hemos asumido que las muestras de los extremos corresponderán a la clase cuya media sea la más cercana, generando probabilidades a 1 o “colas” en esas zonas de la gráfica. Así, evitamos problemas de clasificación en las respuestas al filtro, producidos por una varianza alta en una de las dos clases.

Se observa un alto solape entre la clase homogénea (en rojo) y la clase línea (en azul). Este solape dependerá de la ubicación del umbral. Al haber utilizado un umbral que es más flexible con la clase línea y más estricto con la clase homogénea, el resultado es el que observamos.

La elección de este umbral se ha tomado para obtener más información de píxeles que pueden ser línea, ya que el objetivo final de este TFG depende en gran medida de la fidelidad con la que se clasifican las líneas en la parte inferior de la imagen.

No obstante, se pueden utilizar otros umbrales, ya que la segmentación final dependerá de la combinación de la información que se obtiene de este modelo y de la que se extrae del modelado basado en características de intensidad, que veremos a continuación. Este modelo también requiere de decisiones sobre umbrales para separar las clases, de manera que siendo más flexibles en un modelado y menos en el otro podemos obtener resultados igualmente buenos.

3.2.2 Modelado basado en Características de Intensidad (*Intensity Features Extractor, IFE*)

Es el segundo extractor de parámetros que utilizaremos para clasificar cada píxel de la imagen. Se procesa en paralelo al LMD.

La mayor ventaja de este módulo es el alto nivel de diferenciación entre la clase objeto y pavimento.

Buscamos modelar las tres clases de acuerdo al modelo gaussiano. Esto es, obtener las $P(I_{xy}|X_i)$ o probabilidades de que un píxel tenga una intensidad I_{xy} en función de la clase X_i , donde i puede ser objeto, pavimento o línea. Son funciones de verosimilitud, al igual que en el LMD.

Primero se modelará la probabilidad de pertenencia a la clase pavimento, para más adelante obtener las muestras de las otras dos clases umbralizando. Esta decisión se toma debido a que esta clase es la que tiene más muestras y es la más sencilla de obtener [3].

Para ello, adecuando la ecuación (1) a nuestro caso, se propone:

$$P(I_{xy}|X_{pav}) \cong \exp\left(-\frac{1}{2\sigma_{I,pav}^2} (I_{xy} - \mu_{I,pav})^2\right)$$

donde $P(I_{xy}|X_{pav})$ denota la probabilidad de un píxel de coordenadas x, y , de tener una intensidad I_{xy} sabiendo que la clase X a la que pertenecen dicho píxel es pavimento pav .

Sabiendo esto, para modelar la clase pavimento necesitaremos píxeles exclusivamente de dicha clase. Para eliminar todos aquellos píxeles que no son de pavimento, se utilizan dos métodos combinados:

Primero, sabemos que las regiones de objeto o línea van a estar separadas de las regiones de pavimento por un borde que presentará un alto gradiente en relación a su entorno. Aplicando el detector de Sobel [24] y aplicando posteriormente una dilatación morfológica [25] a los bordes devueltos por el detector, obtendremos una máscara que elimina muchos píxeles que no son de la clase pavimento, como puede observarse en la figura 3.5:

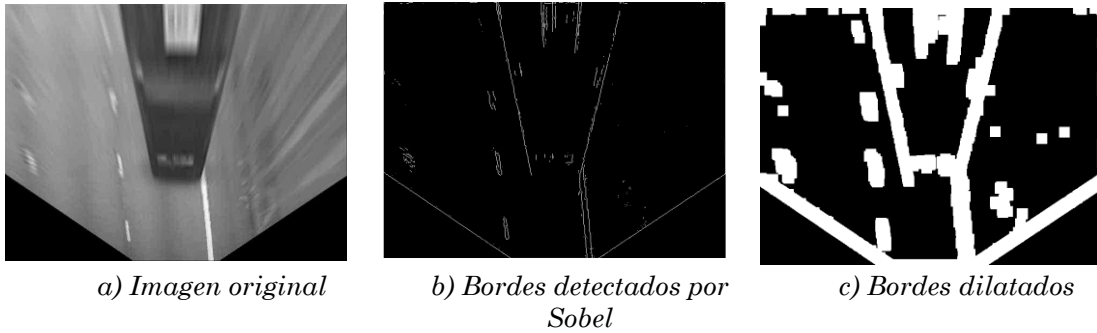


Figura 3.5 Obtención de la máscara por medio del detector de bordes de Sobel y su posterior dilatación morfológica

Adicionalmente, para mejorar esa máscara, se clasifica la imagen en dos clases tomando como umbral el propuesto con el método de Otsu [26], de manera que los píxeles con una intensidad baja serán fácilmente detectados.

Esta propuesta que realizamos en este Trabajo nos es muy útil para no tomar como pavimento la zona inferior del vehículo, que es siempre más oscura debido a la sombra que genera sobre el pavimento y que nos interesa de cara a modelar la clase objeto.

Así, la combinación de ambas máscaras daría como resultado el proceso mostrado en la figura 3.6:



Figura 3.6 Obtención de la máscara del vehículo por Otsu y combinación de los dos métodos

La zona que nos asegurará menor error será la cercana a la parte frontal de nuestro vehículo, recuadrada en la figura 3.7.c, dado que el proceso de obtención de la IPM añade mucho ruido en el resto de la imagen. Las máscaras aplicadas dan un buen resultado para quedarnos únicamente con píxeles de clase pavimento, como podemos observar en la misma imagen.

Con esto, aseguramos de manera fiable que las muestras que utilizaremos (valores distintos de negro en la figura 3.7.c) pertenecen casi exclusivamente a pavimento. Como estamos muestreando la clase pavimento, y no tomando toda su población, los tres pasos anteriores requieren un alto grado de compromiso a la hora de rechazar muestras, ya que hacerlo en exceso puede afectar significativamente a la calidad del modelo.

De las muestras obtenidas extraemos la media y varianza para modelar la clase pavimento. Para las dos clases restantes, será necesario realizar una umbralización.

De acuerdo a una de las propiedades de la función de densidad gaussiana, el 99,74 % de la población se encuentra en el intervalo de $\mu \pm 3\sigma$. Es por ello que en [3] se escogen estos valores como umbral para la clase pavimento, tomando como objeto todas las muestras cuyo valor de intensidad esté por debajo de $\mu_{pav} - 3\sigma_{pav}$ y como línea las de valor por encima de $\mu_{pav} + 3\sigma_{pav}$.

Sin embargo, para mejorar el rendimiento del sistema en el funcionamiento conjunto con el LMD, hemos optado por un umbral adaptativo, que responde a la forma:

$$[\mu_{pav} - \alpha * \sigma_{pav}; \mu_{pav} + \beta * \sigma_{pav}]$$

Para fijar α , tendremos en cuenta el número de píxeles clasificados como objeto en la parte inferior de la imagen, ya que un número alto en esta zona es muy improbable por ser la más cercana al frontal de nuestro vehículo (implicaría prácticamente haber chocado con otro objeto). Si el número de píxeles en la zona de interés tomados como objeto es cercana al 40%⁵, aumentaremos el valor de α a 5, que es el que mejor resultado nos ha proporcionado de forma empírica. Para el resto de valores $\alpha = 3$.

Para fijar el valor de β tendremos en cuenta la respuesta al filtro de líneas, ya que nos ofrece una buena aproximación de la localización de las líneas de carretera. Atendiendo a cuántos píxeles se clasifican como línea en el IFE respecto al LMD, podemos obtener un valor de pérdidas en la clasificación. Si este valor de pérdida supera el 50%⁶, β será igual a 3, como se propone en [3]. Para el resto de los casos, utilizaremos un β igual a 4, ya que, al igual que para α , es el que mejores resultados nos ha dado de forma empírica.

Así, obtenemos las medias y varianzas que modelan las clases objeto y línea, que será la salida real de este bloque. De otra manera, habremos parametrizado las clases cuyas probabilidades serían $P(I_{xy}|X_{pav})$, $P(I_{xy}|X_{obj})$ y $P(I_{xy}|X_{lin})$, siendo las probabilidades de tener una intensidad I_{xy} sabiendo que las muestras que tenemos son de clase pavimento *pav*, objeto *obj* o línea *lin*.

Los modelos que representarían la intensidad de esta manera serían las mostradas en la figura 3.7, una vez que se normalizan las gaussianas para cada píxel de la imagen. Al igual que en el LMD, asumimos que las muestras de los extremos corresponderán a la clase con media más cercana ellas:

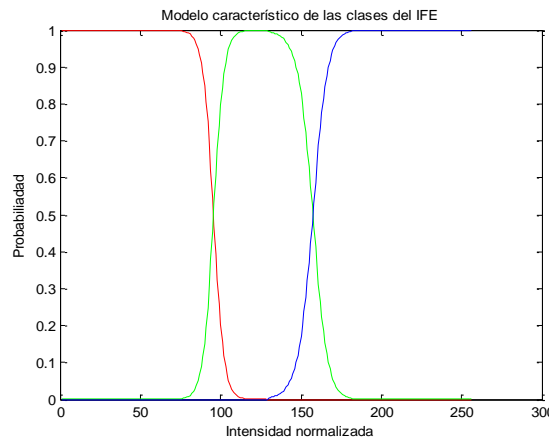


Figura 3.7 Gaussianas obtenidas a partir de los parámetros del IFE, que caracterizarían las tres clases a definir antes de normalizarse. En rojo, la clase objeto; en verde, la clase pavimento; en azul, la clase línea.

⁵ Debido al análisis realizado, este porcentaje de pérdidas es el que mejor funcionaba.

⁶ Igual que la nota al pie [5].

3.3 Refinamiento del modelado mediante el Algoritmo Expectation-Maximitation (EM)

La entrada de este bloque serán las medias y varianzas de cada una de las clases anteriormente obtenidas. El algoritmo *Expectation Maximitation*(EM), es un método iterativo para encontrar una distribución de máxima verosimilitud (explicada en el Anexo 1) de los parámetros que definen una distribución de probabilidad cuando existen datos ocultos o erróneos [27].

Devuelve unas medias y varianzas de las distribuciones que se ajustan lo más fielmente posible probabilísticamente hablando a una población dada.

Podemos aproximar el modelo obtenido a una Mezcla de Gaussianas (*Gaussian Mixture Model*, GMM). De esta manera, podemos utilizar el EM para modelos de este tipo.

La salida del EM serán las medias y varianzas del LMD y del IFE ajustadas a la imagen completa, ya que lo anterior sólo caracterizaba una parte muy filtrada de la imagen.

3.4 Clasificación de la imagen utilizando Máxima Verosimilitud o Máximo a Posteriori

Para poder clasificar nuestra imagen, en este bloque utilizaremos la información que nos proporcionan los modelos obtenidos anteriormente.

Al igual que en el LMD y en el IFE, las clases pueden modelarse muy bien por medio de funciones de densidad de forma gaussiana, por lo que seguimos utilizando este modelo para clasificar los píxeles.

Hay dos principios básicos que nos van a permitir combinar la información obtenida:

- ❖ El primero, la asunción de que la probabilidad de un píxel de pertenecer a una clase considerando el LMD es independiente de la probabilidad de pertenecer a una clase en el IFE. De esta manera, nos será más fácil calcular las probabilidades de un píxel de ser de clase objeto, pavimento o línea a partir de los modelos anteriores.
- ❖ El segundo, que la suma de las probabilidades de cada clase ha de ser uno, es decir, estarán normalizadas. Por ello, la probabilidad de un píxel en el LMD de ser homogéneo más la probabilidad de ser línea es

uno. Igual ocurrirá en el IFE con las probabilidades de ser objeto, pavimento y línea.

De esta manera, con las medias y desviaciones ya obtenidas y por medio de la función de densidad de cada clase, obtendríamos: $P(L_{xy}|X_{hom})$, $P(L_{xy}|X_{lin})$ para el LMD, que serían las probabilidades de tener una respuesta al filtro L_{xy} en caso de ser de la clase homogénea X_{hom} y en caso de ser de la clase línea X_{lin} ; $P(I_{xy}|X_{obj})$, $P(I_{xy}|X_{pav})$, $P(I_{xy}|X_{lin})$ para el IFE, que serían las probabilidades de tener una intensidad I_{xy} en el caso de que la clase sea objeto X_{obj} , pavimento X_{pav} o línea X_{lin} .

Sabiendo que en el LMD la probabilidad de ser objeto es la misma que la de ser pavimento al ser la misma clase que hemos denominado homogénea hom , lo que se escribiría como $P(L_{xy}|X_{obj}) = P(L_{xy}|X_{pav}) = P(L_{xy}|X_{hom})$, podemos escribir de manera compacta que:

$$P(z_{xy}|X_j)|_{j \in \{obj, pav, lin\}} = P(I_{xy}|X_j) * P(L_{xy}|X_j)$$

donde $P(z_{xy}|X_j)$ denota la probabilidad de que una muestra tenga unas características z_{xy} , dependientes de la intensidad y de su posición; sabiendo que esa muestra pertenece a la clase X_j .

3.4.1 Clasificación basada en Máxima Verosimilitud

De acuerdo a la probabilidad obtenida para cada píxel según su clase, $P(z_{xy}|X_j)$, en este bloque realizaremos la asignación de cada píxel a una clase de acuerdo a aquella que presente máxima verosimilitud, método explicado en el Anexo 1.

Esto sería una primera y muy básica clasificación de la imagen, y se utilizará en las pruebas para contrastar las mejoras introducidas por nuestro sistema.

3.4.2 Clasificación basada en Máxima a posteriori

El segundo clasificador de píxeles para la segmentación de la imagen se basa en la teoría de decisión bayesiana, explicada en el Anexo 1. El uso de este tipo de clasificación nos permite introducir conocimiento *a priori* sobre las características de la escena para poder realizar una clasificación más fiable.

En la clasificación según máximo a posteriori, se clasifica según el parámetro z_{xy} que maximiza la ecuación:

$$P(X_j|z_{xy}) = \frac{P(X_j) * P(z_{xy}|X_j)}{\sum_{i=0}^N P(z_{xy}|X_i) * P(X_i)}$$

en la que, de acuerdo con la notación empleada, $P(z_{xy}|X_j)$ denota la probabilidad de que una muestra de posición x, y tenga unas características z_{xy} , dependientes de la intensidad y de la respuesta al filtro; sabiendo que esa muestra pertenece a la clase X_i . De manera análoga, $P(X_j|z_{xy})$ denota la probabilidad de que una muestra sea de la clase X_j , partiendo de que las características de ese píxel son z_{xy} .

Lo importante de esta ecuación es que para obtener la probabilidad *a posteriori*, $P(X_j|z_{xy})$, es necesario disponer de información *a priori*, $P(X_j)$. Debemos ser capaces de obtener la probabilidad de un píxel de pertenecer a una u otra clase sin hacer asunciones sobre su intensidad o su entorno.

3.4.2.1 Generación de *prioris*

La forma de introducir conocimiento previo de la imagen que hemos empleado en este trabajo está basada en el movimiento, para ampliar trabajos relacionados en el Estado del Arte.

La idea de obtener *prioris* en función del movimiento se basa en lo siguiente: si somos capaces de estimar cuánto se está moviendo la carretera por un lado (clases pavimento y línea) y cuánto se están moviendo los vehículos (clase objeto), podremos hacer estimaciones probabilísticas de una imagen a otra proyectando los mapas de probabilidad empleados en la clasificación. Proponemos cuatro casos:

1. Considerando como *priori* el mapa de probabilidades de la imagen anterior (a partir de ahora, hablaremos de “estático”), lo que requiere suponer que el movimiento entre imágenes es muy pequeño.
2. Considerando una gaussiana omnidireccional, cuya varianza en cualquier dirección dependa del módulo del vector de desplazamiento. No se aproximaría al movimiento real aún, pero se ajustaría más que el modelo anterior.
3. Considerando gaussianas que favorecen los dos sentidos posibles del movimiento en cada eje (a partir de ahora, hablaremos de

“bidireccionales”), de acuerdo a las dos direcciones del vector de desplazamiento \vec{v} , \vec{v}_x y \vec{v}_y , por lo que serán simétricas respecto a éstos.

4. Considerando gaussianas que favorecen sólo uno de los dos sentidos del movimiento (a partir de ahora, hablaremos de “direccionales”), que tengan una directividad mayor en el sentido del desplazamiento.

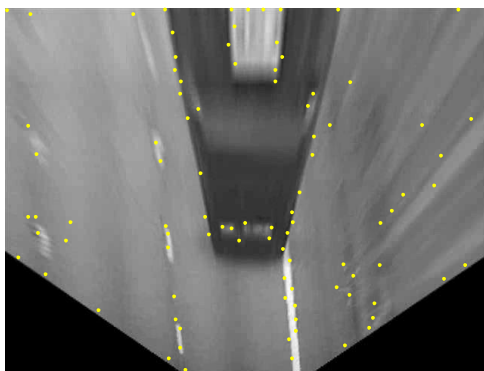
La metodología para la obtención de gaussianas será similar en todos ellos. El uso de gaussianas se debe de nuevo a la facilidad de su modelado. Potenciando direcciones de filtrado de acuerdo a la velocidad, es una manera de proyectar las probabilidades de acuerdo a ésta.

El primer paso será la obtención de vectores de movimiento. Algoritmos como el Block-Matching o el KLT son válidos para este propósito, si bien hemos decidido escoger el segundo para este TFG por su eficacia y mayor velocidad de procesamiento respecto al primero.

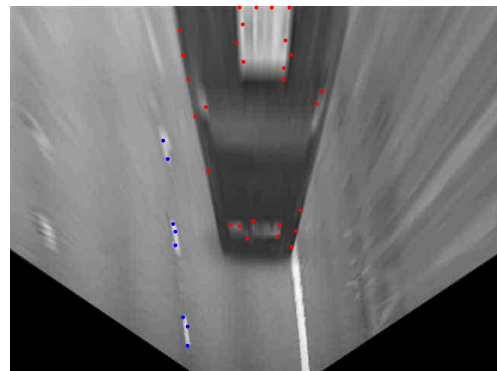
Como viene explicado en el Anexo 2, el algoritmo KLT [28] será capaz de seguir puntos de la imagen en función de su intensidad. Sin embargo, la elección de qué puntos seguir no es trivial.

Para conseguir estos puntos, utilizaremos un detector propuesto por Harris [29], ya que los puntos situados en los bordes con elevado nivel de esquinosidad obtienen un resultado mejor en el seguimiento, por ser más distintivos que los puntos del entorno, reduciendo así el problema de la apertura. Diferenciaremos entre puntos de línea y puntos de objeto para obtener una gaussiana distinta para cada clase.

Finalmente, de todos los puntos que nos devuelve Harris, nos quedaremos sólo con un conjunto escogido por su alta respuesta al detector. El proceso podemos verlo en la figura 3.8:



a) Puntos obtenidos con Harris



b) Puntos útiles después de diferenciar la información. En azul, puntos tomados para línea, en rojo, para objeto

Figura 3.8 Obtención de puntos con los que inicializaremos el KLT

Con los puntos obtenidos que se pueden ver en la figura 3.8.b, utilizamos KLT entre el instante anterior $t-1$ y el instante actual t . Así, obtenemos los vectores de movimiento \vec{v} en función de las coordenadas x , y de la imagen.

La generación del *kernel gaussiano* para el filtrado se realizará de acuerdo a la ecuación de una gaussiana en dos dimensiones:

$$f(x, y) = \exp\left(-\left(\frac{(x - x_o)^2}{2\sigma_x^2}\right) + \left(\frac{(y - y_o)^2}{2\sigma_y^2}\right)\right)$$

donde x_o e y_o serían el centro espacial de la gaussiana, σ_x^2 y σ_y^2 las varianzas en las coordenadas x e y . Desarrollándolo en función del ángulo:

$$f(x, y) = \exp(-(a(x - x_o)^2 - b(x - x_o)(y - y_o) + c(y - y_o)^2))$$

Con:

$$a = \frac{\cos^2 \theta}{2\sigma_x^2} + \frac{\sin^2 \theta}{2\sigma_y^2}$$

$$b = \frac{\sin 2\theta}{4\sigma_x^2} + \frac{\sin 2\theta}{4\sigma_y^2}$$

$$c = \frac{\sin^2 \theta}{2\sigma_x^2} + \frac{\cos^2 \theta}{2\sigma_y^2}$$

Y el ángulo θ se define como $\tan^{-1} \frac{\vec{v}_x}{\vec{v}_y}$, siendo \vec{v}_x y \vec{v}_y las componentes x e y del vector de movimiento \vec{v} .

De esta manera, obtendremos dos gaussianas por cada imagen: una para línea, que se aplicará a los puntos del pavimento por moverse de la misma manera, y otra para objeto. Un ejemplo de los filtros empleados se presenta en la figura 3.9:

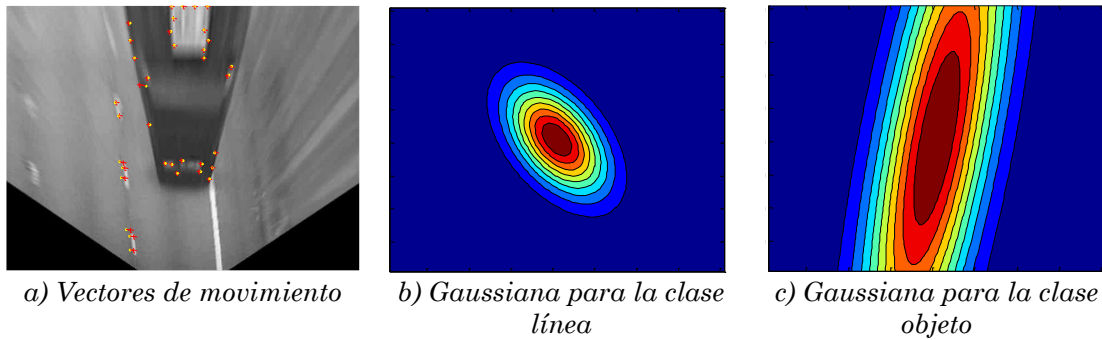
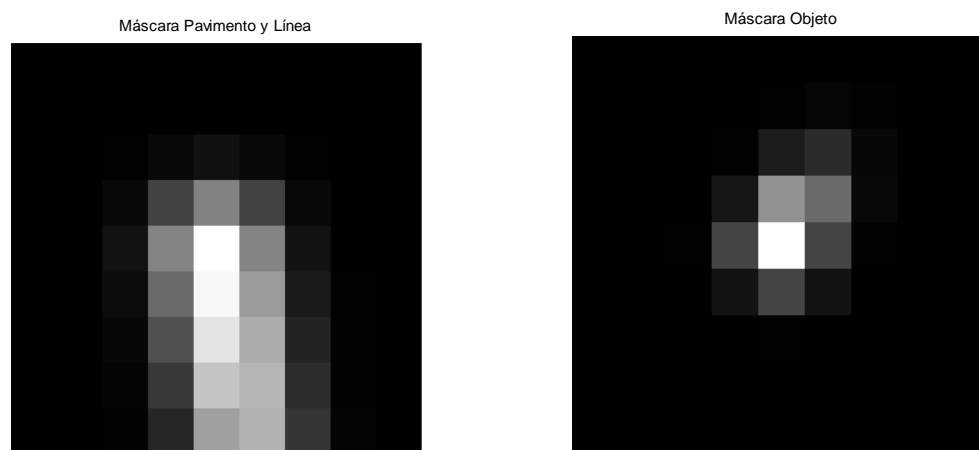


Figura 3.9 Obtención de las gaussianas, previo paso a la obtención de máscaras

Lo que podemos ver en la figura 3.10 sería un ejemplo de gaussianas bidireccionales, es decir, sin tener en cuenta el sentido del movimiento, solo la dirección.

A partir de estudios empíricos, hemos visto que el movimiento máximo medio de un píxel en cada dirección del eje de coordenadas entre una imagen y la siguiente es de dos píxeles. Es decir, un desplazamiento de 2 píxeles en cada sentido supone un área de posible movimiento de 5x5 píxeles⁷.

Para favorecer dos veces el posible desplazamiento, el tamaño que tomaremos para las máscaras gaussianas que filtrarán la imagen de probabilidades tendrá un tamaño 9x9. El resultado para gaussianas direccionales es el mostrado en la figura 3.10:



a) Máscara gaussiana para la clase línea y la clase pavimento b) Máscara gaussiana para la clase objeto

Figura 3.10 Obtención de las máscaras gaussianas

Como está explicado anteriormente, filtrando las probabilidades de la imagen anterior con la máscara gaussiana obtenida, podemos obtener las probabilidades “movidas”, a forma de predicción para la imagen actual.

3.5 Control de segmentación

Este bloque regula la segmentación y estima si ha habido cambios bruscos en la clasificación entre imágenes para controlarlo. Podemos diferenciar tres niveles.

⁷ En el eje vertical, dos hacia arriba, dos hacia abajo y la propia posición del píxel en el centro. En el eje horizontal, dos hacia la derecha, dos hacia la izquierda y el propio píxel en el centro.

En el primer nivel, valoraremos el rendimiento de los umbrales actuales en el LMD y el IFE.

Para la variación del umbral de objeto del IFE, se compara en este bloque la imagen segmentada con el modelo propuesto por el parametrizador. En caso de no cumplir con los requisitos explicados en dicho apartado, se cambiará el umbral.

De igual manera, el cambio del umbral de línea se basará en la comparación de los modelos propuestos por el IFE y el LMD.

En cualquiera de los casos, si se requiere un nuevo cálculo del umbral, se bloquea la salida y se segmenta de nuevo la imagen.

En el segundo, para mejorar la detección de objetos y reducir los errores de clasificación en el interior de los objetos (huecos) que suelen presentar estas regiones, se propone en este trabajo una operación morfológica de cierre. Para poder obtener cada región de objeto por separado, nos hemos basado en un etiquetado a través del algoritmo Blob-Coloring [30].

Por último, en el tercer nivel, hemos adaptado a nuestro sistema la propuesta de [4] en la que se ponderan medias y varianzas del instante actual con el anterior de acuerdo a unos factores de ponderación, para un crecimiento controlado de las clases.

En su lugar, nosotros usamos esta propuesta para verificar que el crecimiento de las clases no está siendo brusco, de manera que:

$$\begin{aligned}\mu_i' &= \lambda_i * \mu_{i,t} + (1 - \lambda_i) * \mu_{i,t-1} \\ \sigma_i' &= \gamma_i * \sigma_{i,t} + (1 - \gamma_i) * \sigma_{i,t-1}\end{aligned}$$

Donde μ_i' , σ_i' son la media y la varianza estimadas de la clase i , $\mu_{i,t}$, $\sigma_{i,t}$ son la media y la varianza de la clase i en el instante t , $\mu_{i,t-1}$, $\sigma_{i,t-1}$ son la media y la varianza de la clase i en el instante $t - 1$, y λ_i , γ_i son los factores de ponderación para controlar la media y la varianza de la clase i .

Considerando $\frac{\mu_{i,t}}{\mu_i'}$ y $\frac{\sigma_{i,t}}{\sigma_i'}$, podríamos tomar decisiones en función del cociente entre los parámetros estimados y los parámetros obtenidos. Si el cociente se aleja de uno de manera considerable, seleccionaremos la imagen anterior como salida del sistema.

3.6 Detección de posición

Para la detección de posición, en este TFG hemos introducido un sistema basado en el conocimiento de la posición de nuestro vehículo en el centro de la imagen y de la posición de las líneas de carretera.

Analizando el histograma por columnas de los píxeles clasificados como línea en una región cuadrada centrada en la zona inferior de la imagen⁸, podemos obtener los máximos locales, que nos darían la ubicación de las líneas.

Para hacerlo más robusto, mejoraremos dicho sistema dilatando las líneas, haciéndolo menos susceptible a la orientación de éstas a la hora de calcular el histograma. Utilizaremos el gradiente espacial en lugar del conteo de píxeles clasificados como línea para calcular el histograma, obteniendo así la posición de los bordes de las líneas.

Se propone un modelo en el que la posición del centro del carril en el que nos encontramos se mide en píxeles respecto al centro de la imagen (que sería el centro de nuestro vehículo). Así, posiciones positivas implican encontrarnos a la izquierda del centro del carril, mientras que una posición negativa significaría encontrarnos a la derecha. Por tanto, el cero es estar en el centro exacto del carril.

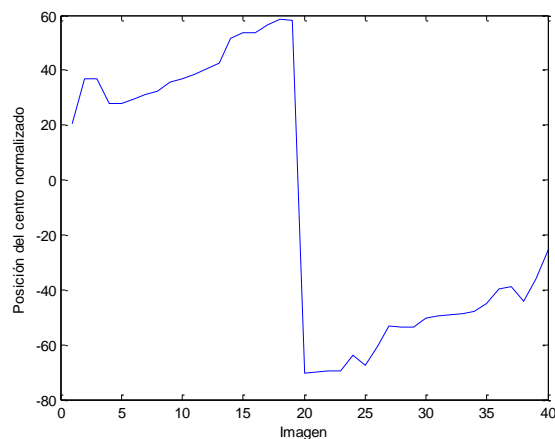


Figura 3.11 Posición del centro del carril respecto al vehículo

Como un cambio de carril supondrá un cambio de signo en la gráfica, podremos utilizar este dato para activar una señal de control que nos avise de dicho suceso. Esto puede verse en la figura 3.11, en la que de la imagen 19 a la 20 se ha considerado que el cambio de carril se ha efectuado. De acuerdo a lo explicado, sería un cambio de carril a la izquierda.

⁸ Recordar que esta área es el que se encuentra frente al vehículo, tanto en el dominio espacial como en el transformado.

Capítulo 4: Pruebas y resultados

4.1 Introducción

El objetivo de las pruebas es realizar una evaluación objetiva del sistema. Para ello, compararemos cuatro versiones del sistema, adelantadas en el apartado 3.1: máxima verosimilitud sin EM, máxima verosimilitud con EM, máximo a posteriori con *prioris* estáticos sin EM y máximo a posteriori con EM.

Para las versiones que utilizan Bayes y EM conjuntamente, se evaluará la opción de utilizar *prioris* basados en movimiento mediante el filtrado de los mapas de probabilidad con máscaras gaussianas bidireccionales o *prioris* basados en movimiento filtrando los mapas de probabilidad con gaussianas direccionales, ya que son las dos opciones más complejas del sistema.

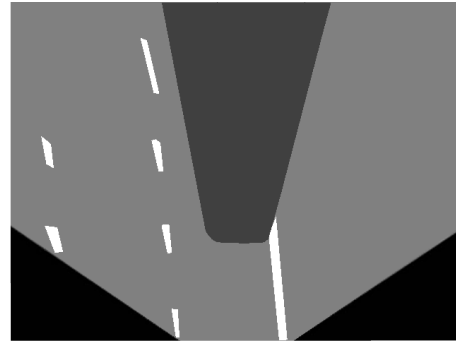
En primer lugar, se explicarán las medidas propuestas para analizar la segmentación. Después, se explicarán los resultados obtenidos con dichas pruebas.

Por último, se analizará subjetivamente el sistema comparando las imágenes segmentadas con las imágenes de referencia, y analizaremos el comportamiento del cambio de carril.

Para la evaluación objetiva se utilizarán unas imágenes etiquetadas de referencia (*ground-truth*). Podemos ver un ejemplo de este tipo de imágenes en la figura 4.1. Como se puede observar, la imagen derecha está etiquetada de acuerdo a tres clases, en blanco las líneas, en gris claro el pavimento y en gris oscuro el objeto:



a) Imagen IPM original



b) Segmentación de referencia ground truth

Figura 4.1 Imagen original y su ground truth

4.2 Medidas

Para cada clase se plantean una serie de medidas [4] que permiten analizar el rendimiento de los algoritmos propuestos. Se tendrá en cuenta en cada caso la información potencial que queremos extraer de cada una de las clases.

4.2.1 Clase Pavimento

Al no tener una forma geométrica precisa, podemos evaluar la precisión en la clasificación de los puntos pertenecientes al pavimento. Para ello, se utilizarán los indicadores *F-Score*, precisión y acierto, definidos como:

- TP o *True Positive*, píxeles clasificados correctamente como pertenecientes a su clase (pavimento en este caso).
- FP o *False Positive*, píxeles clasificados incorrectamente como pertenecientes a una clase (pavimento en este caso).
- FN o *False Negative*, píxeles clasificados como no pertenecientes a una clase (pavimento) cuando en realidad sí pertenecen a ella.

A partir de estas definiciones se plantean:

- Precision o *precision*, que se representa como $Precision(X_i) = \frac{TP}{TP+FP}$, donde X_i es la clase para la que se define. Da un valor entre 0 y 1 indicando el grado de la precisión con el que estamos clasificando inter-clase.

-Acierto o *recall*, que se escribe como $Recall(X_i) = \frac{TP}{TP+FN}$, donde X_i es la clase para la que se define. Da un valor entre 0 y 1 indicando el grado de la precisión con el que estamos clasificando intra-clase.

Por último, en base a estas dos medidas, se define el *F-Score*:

-*F-Score*, medida armónica entre la precisión y el acierto, que puede escribirse como:

$$Fscore = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2 * Precision * Recall}{Precision + Recall}$$

Valores cercanos a 1 significará que las clasificaciones TP estarán siendo muy altas en comparación a FP y FN.

Por una parte, analizamos las puntuaciones *F-Score* obtenidas para los métodos de máxima verosimilitud sin EM, máxima verosimilitud con EM y máximo a posteriori con *prioris* estáticos sin EM en toda la imagen, ya que serían los métodos más simples del sistema, y que servirán en el caso de que las mejoras no dieran resuemosltado.

Después, evaluar la opción de utilizar *prioris* basados en movimiento mediante el uso de gaussianas bidireccionales o *prioris* basados en movimiento utilizando gaussianas direccionales, ya que son las dos opciones más complejas del sistema.

Para ver su rendimiento, lo mediremos con el mejor de los casos de la primera comparativa.

4.2.2 Clase Objeto

Lo que nos interesa de la segmentación será extraer regiones que nos sirvan como hipótesis de ser una zona potencial para obtener clase objeto, así como regiones de líneas de carretera.

La verificación, que podría hacerse en base a las pruebas propuestas, dependerán principalmente del estudio de esas regiones.

Como puede verse en las imágenes IPM, las zonas que contienen vehículo, que son las que realmente nos interesan de la clase objeto, son zonas trapezoidales con base estrecha en la parte baja de la imagen. Esto se debe a la operativa para obtener dichas imágenes IPM. Puede verse en la figura 4.2 (en rojo la región correspondiente al vehículo):

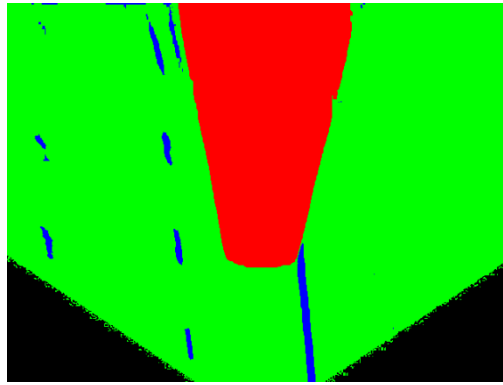


Figura 4.2 Ejemplo de la geometría de un vehículo debido a la transformada IPM

Por ello, dos métodos serán los que utilizaremos para estudiar la bondad de nuestro clasificador con vehículos: la medida de los ángulos de los laterales del vehículo y la distancia en píxeles a la que se encuentra la base de éste (extremo inferior de la sombra correspondiente a la parte más próxima del vehículo) respecto a la base de la imagen.

La forma de obtener estos ángulos y distancias será por medio de la transformada de Hough [31], que inicializaremos con los bordes resultado de aplicar un detector de Canny [32] a la imagen.

De manera similar al caso de la clase pavimento, evaluaremos, por una parte, los ángulos obtenidos para los métodos de máxima verosimilitud sin EM, máxima verosimilitud con EM y máxima a posteriori con *prioris* estáticos sin EM en toda la imagen, y después, compararemos los métodos de máximo a posteriori con EM y *prioris* gaussianos bidireccionales y con *prioris* gaussianos direccionales.

En ambas comparativas utilizaremos de referencia las medidas obtenidas por las imágenes *ground-truth*.

4.2.3 Clase Línea

Para la clase de línea acudimos a su geometría y disposición en la carretera. Por lo general, la clase línea estará dividida en varias regiones rodeadas de otra clase no línea.

Nos interesa que la información de los carriles sea fácilmente extraíble, de manera que se evaluará el número de regiones detectadas como línea y su orientación. Para ello, como en todas las pruebas propuestas, se hará uso de las imágenes de referencia *ground truth*, centrando el análisis en la región central-inferior de la imagen ya que es de donde utilizaremos la información de líneas para establecer los límites del carril.

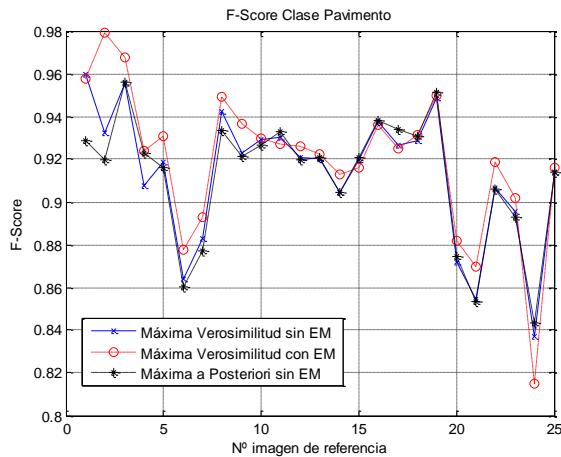
La orientación de las líneas se evaluará empleando nuevamente transformada de Hough, y para la determinación del número de segmentos de línea correctamente detectados se utilizará el algoritmo Blob-Coloring.

Para las comparativas, haremos de manera semejante a la clase objeto. Por una parte, los métodos de máxima verosimilitud sin EM, máxima verosimilitud con EM y máxima a posteriori con *prioris* basados en mapas de probabilidad estáticos sin EM, y después, evaluaremos la opción de utilizar máxima a posteriori con EM y *prioris* basados en movimiento mediante el uso de gaussianas bidireccionales o *prioris* basados en movimiento utilizando gaussianas direccionales.

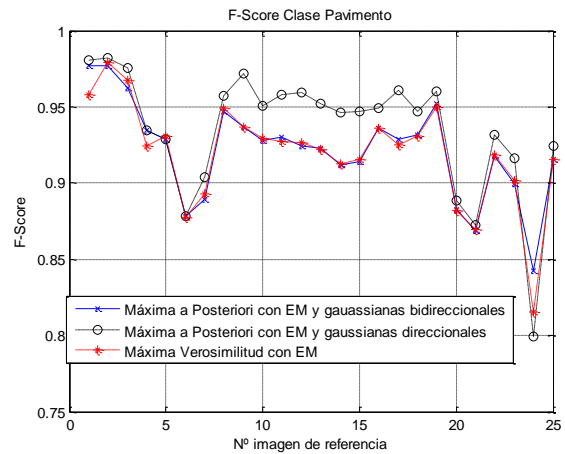
4.3 Resultados

A continuación, se muestran los resultados obtenidos con su análisis, de acuerdo a las pruebas realizadas.

4.3.1 Clase Pavimento



a) *F-Score* clase pavimento para Máxima Verosimilitud con y sin EM y Máximo a Posteriori sin EM con *prioris* basados en mapas de probabilidad estáticos



b) *F-Score* clase pavimento para Máximo a Posteriori con EM y *prioris* basados en mapas de probabilidad obtenidos mediante el uso de gaussianas bidireccionales o direccionales

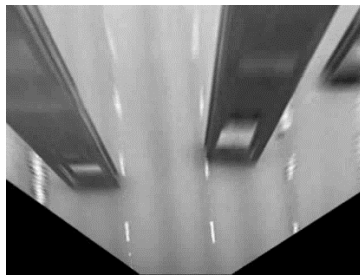
Figura 4.3 Puntuaciones *F-Score* clase pavimento

En la figura 4.3 se muestran las puntuaciones *F-Score* de la clase pavimento: en la imagen 4.3.a se presentan los resultados empleando de máxima verosimilitud (ML) con y sin los parámetros refinados por el EM, así como empleando máximo a posteriori (MAP) sin EM y con *prioris* basados en mapas

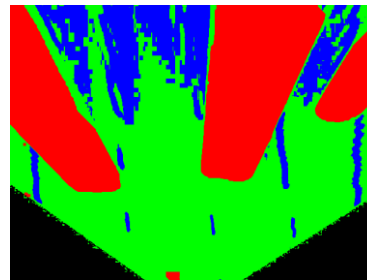
de probabilidad estáticos (en la leyenda, Máxima a Posteriori sin EM); y en la imagen 4.3.b comparamos los resultados empleando MAP con EM y empleando α prioris calculados filtrando los mapas de probabilidad en el instante anterior mediante gaussianas bidireccionales (en la leyenda de la imagen, Máxima a posteriori con EM y gaussianas bidireccionales) y direccionales como *prioris* (en la leyenda, Máxima a posteriori con EM y gaussianas direccionales), enfrentado al mejor de los resultados de la comparativa anterior.

Atendiendo a la imagen 3.4.a, observamos que los resultados son muy similares en todos los casos, con un buen rendimiento siempre por encima del 0,9 excepto en las imágenes de referencia 20, 21, 23 y 24: en las tres primeras esto se debe a que, a pesar de los mecanismos de control que se han adoptado, se han perdido muchos píxeles de línea a favor de píxeles de pavimento, de manera que se dan muchos falsos positivos que no tienen lugar en el resto de la secuencia de prueba.

En la imagen 24 tiene lugar otro fenómeno distinto, que es un pico de ruido en la parte superior de la imagen, clareándola. Esto se ve en la figura 4.4:



a) Imagen ruidosa original



b) Imagen ruidosa segmentada

Figura 4.4 Imagen original e imagen segmentada para el peor resultado *F-Score* obtenido

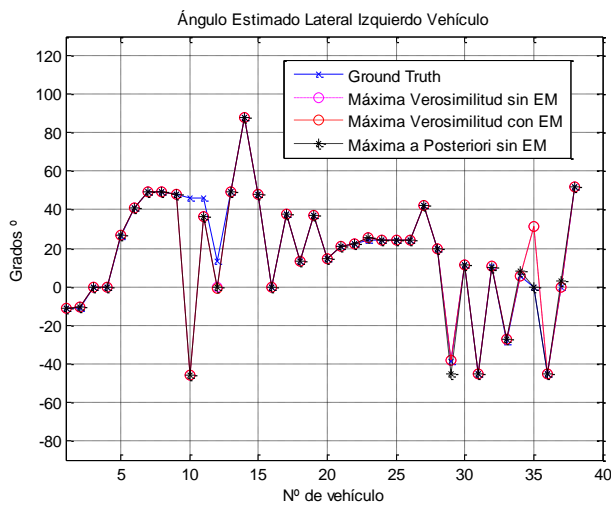
A pesar de las medidas utilizadas, el ruido crece de manera progresiva producto de la transformada IPM, afectando especialmente a la parte superior de la imagen. En consecuencia, se producen errores significativos en la detección del pavimento en esa zona: las medidas tomadas en la zona de interés para establecer el modelo no representan adecuadamente la variabilidad existente en la parte superior.

Se ha tomado el método de máxima verosimilitud con refinado de parámetros con EM como referencia para la figura 3.4.b por ser el que mayor media global *F-Score* presentaba en la comparativa anterior.

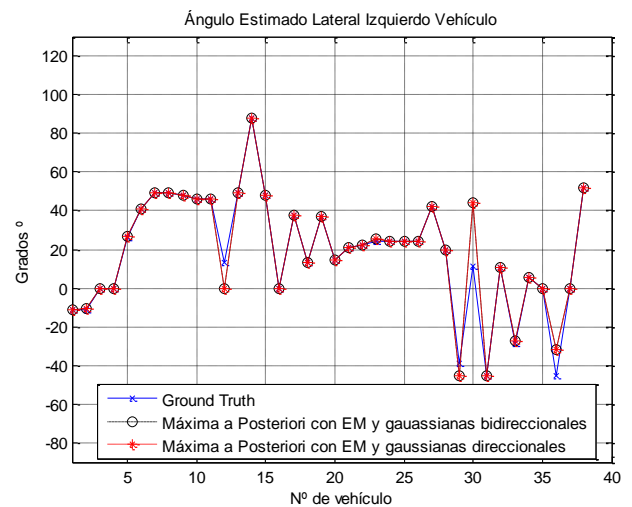
Entre la obtención de *prioris* por medio de gaussianas direccionales o bidireccionales para filtrar el mapa de probabilidad de la imagen anterior, obtienen mejor rendimiento las direccionales de forma bastante clara. Este resultado es extensible a la totalidad de los métodos.

4.3.2 Clase Objeto

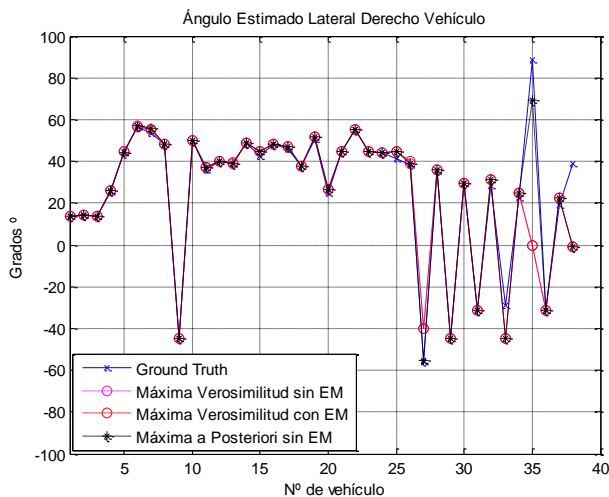
La figura 4.4 muestra los ángulos que parametrizan los laterales de los vehículos. En el eje horizontal de la gráfica, se enumeran los vehículos detectados en total en la secuencia de referencia, tomando como vehículos distintos a aquellos en imágenes diferentes y a aquellos en regiones inconexas de una misma imagen. En el eje vertical de la gráfica, se representan los ángulos de las detecciones, siendo 0° una detección perpendicular al eje horizontal de la imagen.



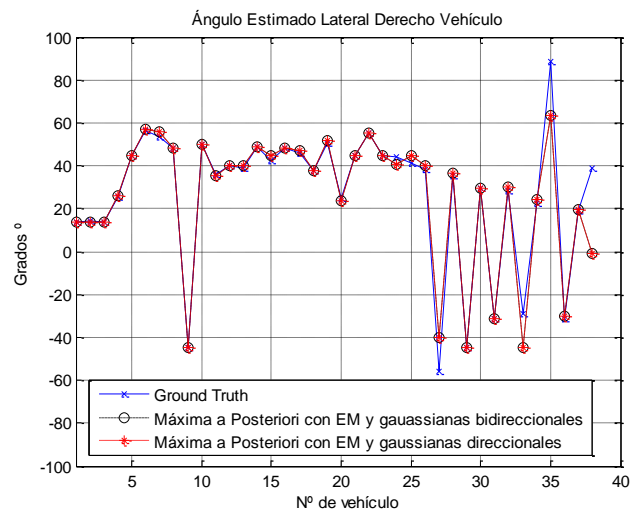
a) Ángulo de la recta que parametriza al lateral izquierdo de los vehículos detectados para Máxima Verosimilitud con y sin EM con prioris basados en mapas de probabilidad estáticos



b) Ángulo de la recta que parametriza al lateral izquierdo de los vehículos detectados para Máxima a Posteriori con EM y prioris basados en mapas de probabilidad obtenidos mediante el uso de gaussianas bidireccionales o direccionales



c) Ángulo de la recta que parametriza al lateral derecho de los vehículos detectados para Máxima Verosimilitud con y sin EM con prioris basados en mapas de probabilidad estáticos



d) Ángulo de la recta que parametriza al lateral derecho de los vehículos detectados para Máxima a Posteriori y prioris basados en mapas de probabilidad obtenidos mediante el uso de gaussianas bidireccionales o direccionales

Figura 4.5 Ángulos que parametrizan los laterales de los vehículos

La primera conclusión que podemos extraer es que todos los métodos tienen un alto rendimiento en casi toda la secuencia. Esto se debe en gran medida a los mecanismos de control que hemos introducido, aplicados a todos métodos. Principalmente, las operaciones morfológicas introducidas dan una buena geometría al vehículo, como puede observarse en la figura 4.6, de manera que el ángulo de los laterales queda bien definido y no tiene mucha dependencia del método utilizado.

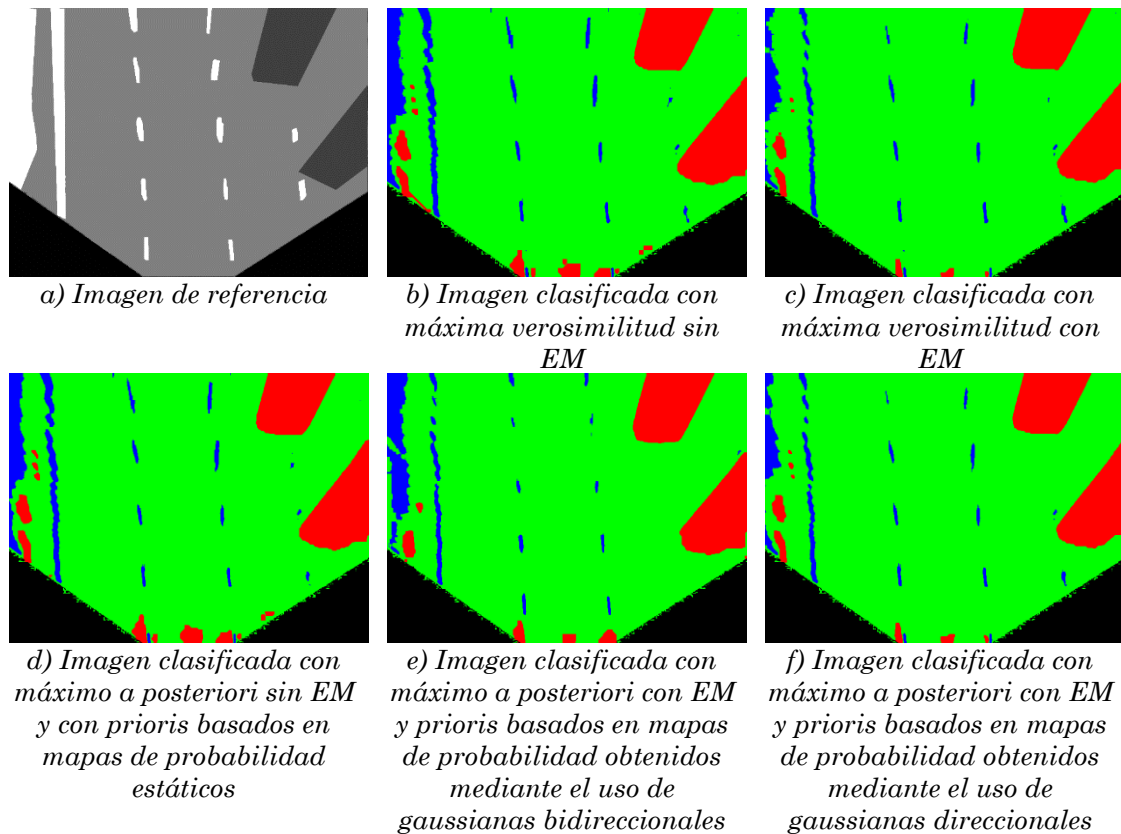


Figura 4.6 Ejemplo de una imagen de referencia a partir de la que comparamos las imágenes segmentadas para cada uno de los métodos. Nótese la semejanza en la geometría del vehículo en todos los casos

En segundo lugar, atendiendo al valor de los ángulos los laterales del vehículo, y significando un valor positivo del ángulo una pendiente positiva y un valor negativo del ángulo una pendiente negativa, llama la atención los cambios repentinos que hay de valor y de signo de un vehículo a otro.

Esto se debe a tres razones. La primera, que la transformada IPM, como ya hemos visto, deforma los vehículos dependiendo de la ubicación éste en la imagen. Vehículos a la derecha tendrán pendientes laterales positivas, mientras que vehículos a la izquierda tendrán pendientes laterales negativas. Así, un vehículo centrado tendría distinto signo en la pendiente de cada uno de sus laterales.

La segunda razón es el contexto de la secuencia. Al inicio, sólo tenemos un vehículo delante, centrado, con un signo de pendiente distinto en cada lateral. Después, adelantamos este vehículo por nuestra izquierda. Al quedar a la derecha el vehículo adelantado, los ángulos que caracterizan sus laterales pasan a ser ambos positivos.

La tercera razón es la que explica los saltos bruscos en la gráfica, y también está relacionada con el contexto. Cuando hay más de un vehículo en la imagen, puede ocurrir que uno esté a la izquierda y otro a la derecha, teniendo cada uno signos distintos en las pendientes de sus laterales. Como la gráfica está en función del conteo de vehículos a lo largo de la secuencia, si en dos imágenes sucesivas tenemos dos o más vehículos en lados distintos de la imagen, habrá un salto seguro en algún momento. Es lo que ocurre claramente a partir de la imagen de referencia 24.

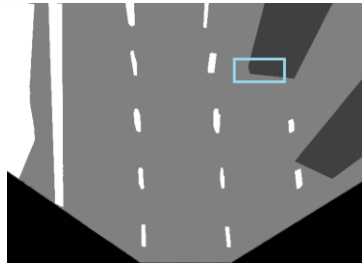
Todo esto se muestra con algunos ejemplos de la secuencia en la figura 4.7:



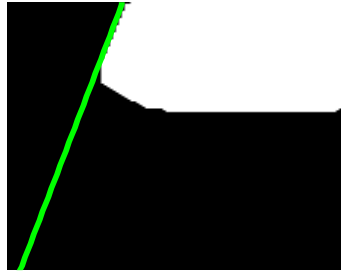
Figura 4.7 Imágenes obtenidas de la secuencia de referencia por orden cronológico

Los fallos producidos en la secuencia, como puede ser las imágenes 10 o 30 para el lateral izquierdo o las imágenes 35 y 38 del lateral derecho, son producidas en su mayoría por objetos pequeños, ya que cambios poco significativos en el borde de este tipo de regiones, pueden dar lugar a ángulos muy diferentes.

Un ejemplo se ve en la figura 4.8, en la que un vehículo en la zona superior da malos resultados a la hora de estimar el ángulo, cuando la segmentación es buena. Se corresponde a la imagen de referencia número 10:



a) Imagen de referencia.
Recuadrado en azul, la zona del vehículo ampliada en b) y c)



b) Pendiente obtenida a través de la transformada de Hough para el vehículo en el ground-truth

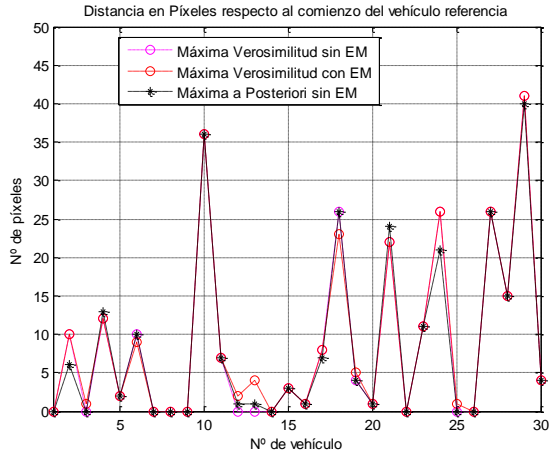


c) Pendiente obtenida a través de la transformada de Hough para el vehículo en la imagen segmentada

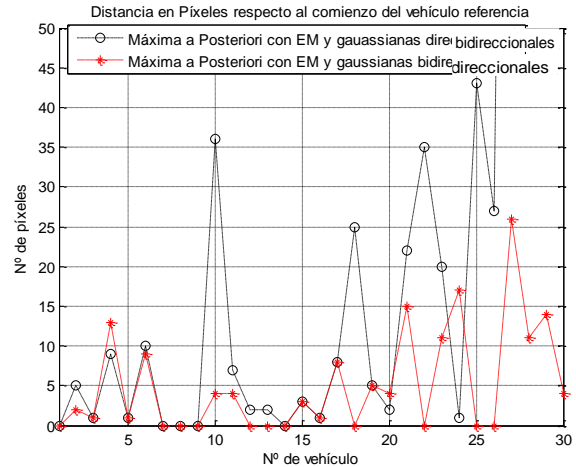
Figura 4.8 Diferencias en el ángulo obtenido para las rectas que caracterizan los laterales, ampliado en la zona que produce el cambio de orientación

Por último, los picos tienen lugar con laterales muy verticales, lo cual puede suceder en momentos puntuales. Puede verse en la figura 4.8.c mostrada anteriormente.

En cuanto a las distancias, mediremos el error cometido como la diferencia en píxeles entre la distancia a la que se encuentra el vehículo detectado del frontal de nuestro vehículo en la imagen *ground-truth* y la misma distancia obtenida en la imagen segmentada. Para los vehículos de toda la imagen, se observan los resultados en la imagen 4.9:



a) Diferencia en píxeles entre los vehículos detectados y los de referencia para Máxima Verosimilitud con y sin EM y Máxima a Posteriori con EM y prioris basados en mapas de probabilidad estáticos



b) Diferencia en píxeles entre los vehículos detectados y los de referencia para Máxima a Posteriori y prioris basados en mapas de probabilidad obtenidos mediante el uso de gaussianas bidireccionales o direccionales

Figura 4.9 Diferencias en píxeles entre el comienzo de los vehículos detectados y los de referencia para toda la imagen

Los errores altos de clasificación se deben en su mayoría a regiones pequeñas y en las zonas influenciadas por el ruido que introduce la transformada IPM. Parte de estos vehículos se pierden en la detección, aunque su geometría suele estar siempre bien definida.

Sin embargo, si restringimos la zona de aplicación a la región de interés (central-inferior de la imagen), los resultados mejoran considerablemente, como puede observarse en la figura 4.10:

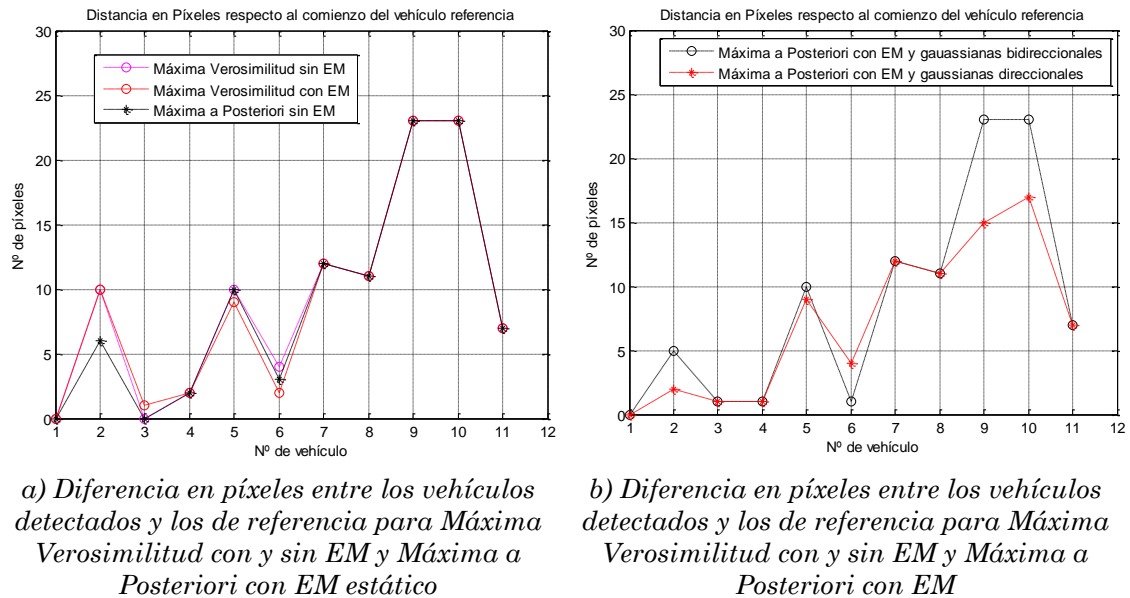


Figura 4.10 Diferencias en píxeles entre el comienzo de los vehículos detectados y los de referencia para la región de interés

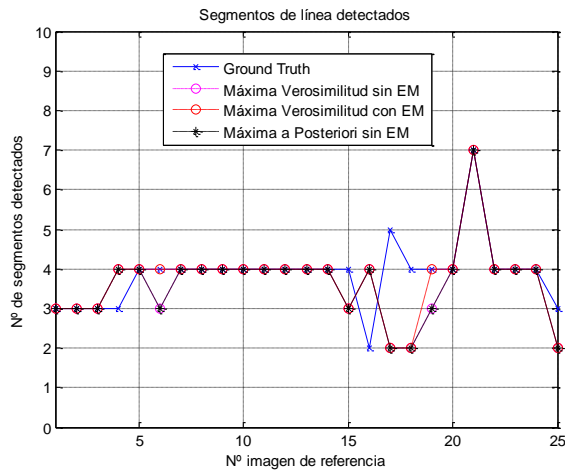
Para 11 detecciones de vehículo, sólo dos de ellas están por encima de los 15 píxeles de error, llegando a los 22-23 píxeles de error, que es aceptable. Se observa también que, ya sea para toda la imagen o para la región de interés, el método de máxima a posteriori basado en *prioris* obtenidos a través del filtrado de los mapas de probabilidad de la imagen anterior con una máscara gaussiana direccional es el que mejores resultados da con diferencia.

Claramente, no utilizar el EM y algún tipo de información de movimiento para los *prioris* hace que el sistema sea muy poco robusto y propenso a errores. Esto disminuye muy significativamente cuando se emplean *prioris* teniendo en cuenta direccionalmente el movimiento.

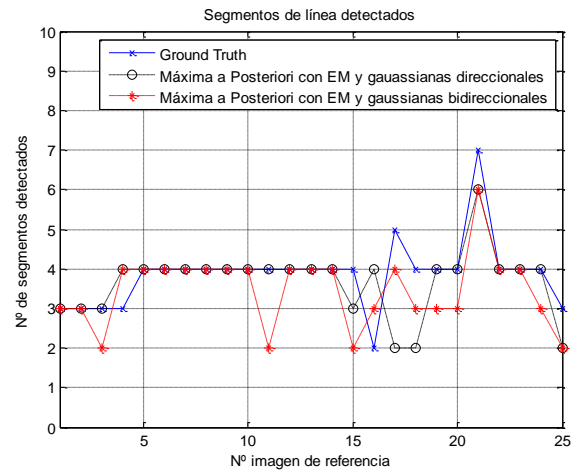
4.3.3 Clase Línea

Para esta clase se han analizado las líneas de carril por las que circula el vehículo sólo en la zona de interés (central-inferior) de la imagen, si bien en los cambios de carril se tienen en consideración carriles adyacentes al entrar en la zona de interés.

En la figura 4.11 se pueden observar los segmentos de línea detectados en el área de interés:



a) Segmentos de línea detectados para Máxima Verosimilitud con y sin EM y Máxima a Posteriori con EM estático

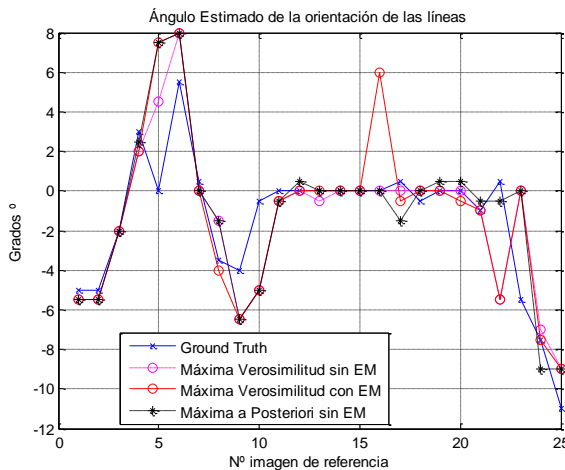


b) Segmentos de línea detectados para Máxima Verosimilitud con y sin EM y Máxima a Posteriori con EM

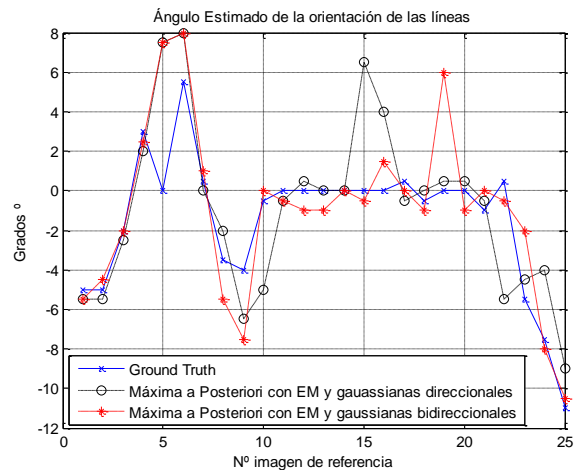
Figura 4.11 Segmentos de línea detectados entre los cinco sistemas y la imagen de referencia

De nuevo, se ve mucha similitud entre la utilización de un método u otro. Los picos de error, como la imagen de referencia 17 o la 19, se deben a pérdidas de información de línea.

La otra medida propuesta, los ángulos de orientación de las líneas, se muestran en la figura 4.12:



a) Ángulos de orientación de las líneas de carretera para Máxima Verosimilitud con y sin EM y Máxima a Posteriori con EM estático



b) Ángulos de orientación de las líneas de carretera para Máxima Verosimilitud con y sin EM y Máxima a Posteriori con EM

Figura 4.12 Ángulos que parametrizan la orientación de las líneas de carretera

Los ángulos de error son muy pequeños, entre los -11° y 8° en los peores casos. La estimación del ángulo de las líneas es muy fiable ya que los bordes que obtenemos son muy directivos.

Por ello, para la detección de líneas, y como conclusión a las dos pruebas, no habría un método claro que se pueda proponer como el mejor. La orientación

que obtienen es generalmente buena en todos los casos, al igual que el número de regiones obtenidas como línea que realmente lo son.

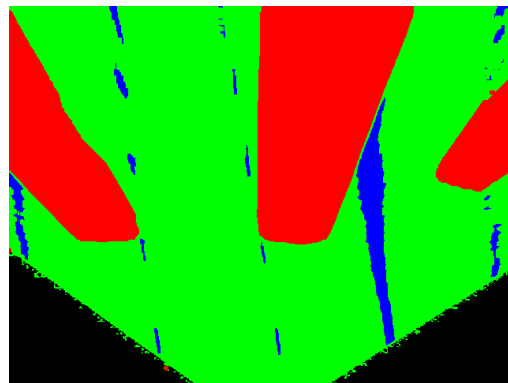
4.3.4 Análisis subjetivo

El análisis subjetivo lo hacemos en función a las secuencias que hemos obtenido durante todo el trabajo de investigación, mostradas a lo largo de este Trabajo de Fin de Grado.

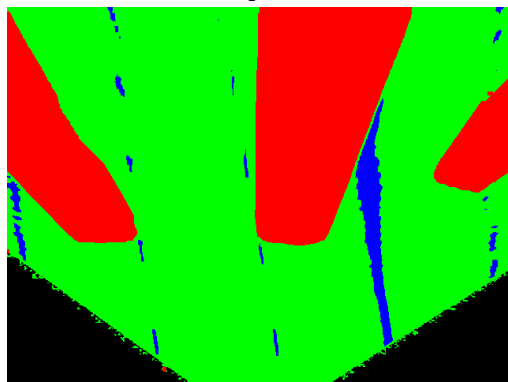
Para los resultados obtenidos, el mejor método dependerá de la aplicación. Por ejemplo, el método de gaussianas direccionales con EM elimina prácticamente todo el ruido de la imagen favoreciendo las clases pavimento y objeto, pero elimina mucha información de la clase línea. Esto puede verse en la figura 4.13, fijándonos en las líneas de la zona inferior de la imagen:



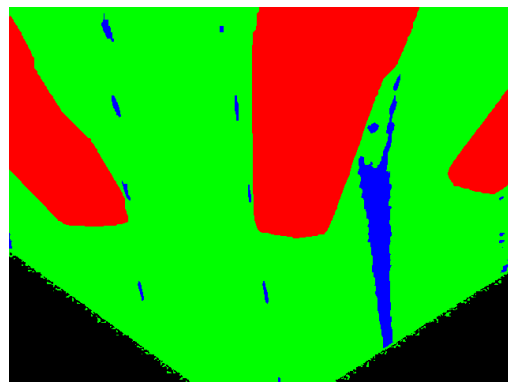
a) Imagen IPM 25 de la secuencia utilizada en las pruebas



b) Imagen 25 segmentada utilizando máxima verosimilitud sin EM



c) Imagen 25 segmentada utilizando máxima verosimilitud con EM



d) Imagen 25 segmentada utilizando máxima a posteriori con EM y prioris basados en mapas de probabilidad obtenidos mediante el uso de gaussianas bidireccionales o direccionales

Figura 4.13 Ejemplo de una imagen IPM segmentada con diversos métodos propuestos

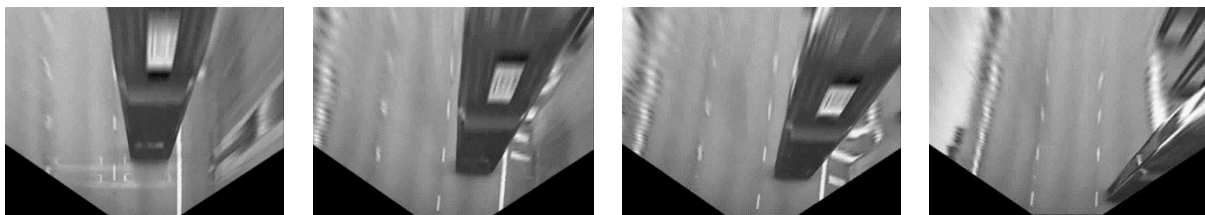
El refinamiento EM, por su parte, mejora la clasificación significativamente, lo que lo hace indispensable sea cual sea la metodología que adoptemos en el

sistema. El uso exclusivo de este método proporciona buenos resultados para la clase línea. También esto se debe a los mecanismos de control introducidos, que mejoran la segmentación mucho.

De forma general, los cambios de iluminación no suelen cambiar nuestro sistema, gracias a los mecanismos de control propuestos que hemos introducido. Este suele ser uno de los mayores problemas en este tipo de sistemas.

Por último, la detección del centro de carril funciona de manera correcta, detectando los cambios de carril como se ha mencionado en el apartado 3.7. Para medirlo con rigor, hemos escogido un rango de imágenes en el que se produce un cambio de carril, ya que determinar el momento exacto en el que se pasa de un carril a otro no es sencillo. Así, si nuestro sistema detecta un cambio de carril en dicho rango, estará funcionando satisfactoriamente.

Concretamente, de las imágenes 61 a 103 hay un cambio de carril, tal y como se ejemplifica en la figura 4.14:



a) Imagen IPM 61 b) Imagen IPM 75 c) Imagen IPM 94 d) Imagen IPM 103

Figura 4.14 Ejemplos de la secuencia de referencia utilizada para la detección del cambio de carril

Podemos fijarnos en las líneas de carretera en las imágenes de la figura 4.7 para verlo mejor. De acuerdo con el modelo propuesto, y siendo el 1 de la gráfica la primera imagen de la secuencia (en este caso, la 47), se muestra la gráfica que describe la posición del vehículo a lo largo del carril en la figura 4.8:

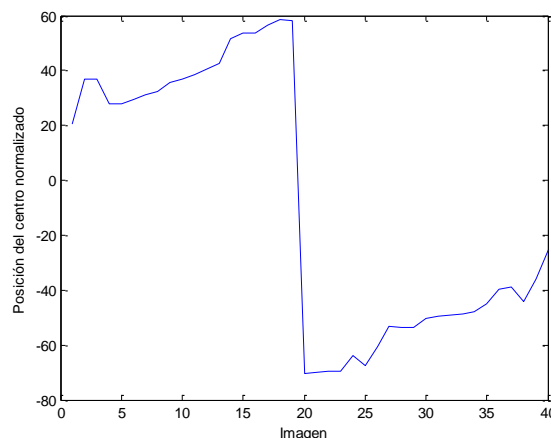


Figura 4.15 Posición del centro del carril respecto al vehículo

Podemos concluir que el funcionamiento del detector de posición es correcto.

Capítulo 5: Conclusiones y trabajo futuro

El objetivo del sistema es detectar la posición de nuestro vehículo respecto al centro del carril en el que nos encontramos y cambios de carril. El análisis subjetivo arroja buenos resultados de acuerdo a la mejor de las segmentaciones presentadas para la detección de líneas en la zona de interés, que sería el sistema con máxima verosimilitud y refinamiento de datos con el algoritmo EM.

Transversalmente, hemos presentado formas de segmentación y estrategias para hacerla más robusta, observando que, en función del problema que deseemos resolver, utilizaremos uno u otro método.

Principalmente, hemos propuesto un sistema de clasificación de la imagen basado en la estadística bayesiana, con *prioris* estimados en función del movimiento. También hemos desarrollado mecanismos de control para hacer más fiable la segmentación de la imagen y evitar cambios bruscos en la clasificación entre una imagen y la siguiente. Estos mecanismos de control han sido fundamentales para poder obtener buenos resultados con métodos más sencillos.

Hemos observado los problemas que añade el ruido de la transformada IPM, y por último hemos añadido un estimador de cambios de carril.

Las formas de mejorar el presente trabajo serían las siguientes:

- ❖ Introducción de una clase desconocido que permitiera modelar la clasificación en zonas con probabilidad baja para las clases propuestas.
- ❖ Utilizar la información del movimiento de manera que no perjudique la clase línea.
- ❖ No realizar una estimación de las distribuciones en todas las imágenes para aliviar el coste computacional.

Las implementaciones principales a partir del presente trabajo son:

- ❖ Integrar la transformada IPM en el sistema y poder trabajar en ambos espacios (real y transformado) para hacer el sistema más robusto.
- ❖ Verificar las regiones generadas como hipótesis.
- ❖ Utilización del filtro Kalman para la estimación de la posición de nuestro vehículo en el carril de una manera más fiable.

Referencias

De acuerdo a la norma IEEE para referencias:

- [1] F. Bernuy, J. Ruiz del Solar, I. Parra y P. Vallejos, «Adaptive and real-time unpaved road segmentation using color histograms and RANSAC,» de *9th IEEE International Conference on Control and Automation (ICCA)*, Santiago, 2011.
- [2] W. H. Organization, «Global status report on road safety 2015,» 2015.
- [3] M. Nieto, J. Arróspide y L. Salgado, «Road environment modeling using robust perspective analysis and recursive Bayesian segmentation,» *Machine Vision and Applications*, vol. 22, n° 6, pp. 927-945, 2011.
- [4] J. I. Bravo Pérez-Villar, «Detección de vehículos mediante análisis de imágenes,» Universidad Autónoma de Madrid, Escuela Politécnica Superior, Madrid, 2015.
- [5] Z. Sun, G. Bebis y R. Miller, «On-road vehicle detection: a review,» *IEEE Transactions on Pattern Analysis and Machine Intelligence* , vol. 28, n° 5, pp. 694-711, 2006.
- [6] J. P. González y Ü. Özgüner, «Lane detection using histogram-based segmentation and decision trees,» *Intelligent Transportation Systems, 2000*, pp. 346-351, 2000.
- [7] L. Dirk, M. Mekhaïel, J. Dickmann y K. C. J. Dietmayer, «Pedestrian detection with thermopiles using an occupancy grid,» *Intelligent Transportation Systems, 2004*, pp. 1063-1068, 2004.
- [8] E. Bas, A. M. Tekalp y F. S. Salman, «Automatic Vehicle Counting from Video for Traffic Flow Analysis,» *2007 IEEE Intelligent Vehicles Symposium* , pp. 392-397, 2007.

- [9] M. Bertozzi, A. Broggi, A. Fascioli y S. Nichele, «Stereo vision-based vehicle detection,» *Intelligent Vehicles Symposium, 2000. IV 2000*, pp. 39-44, 2000.
- [10] M. Bertozzi y A. Broggi, «GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection,» *IEEE Transactions on Image Processing*, pp. 62-81, 1998.
- [11] J. M. Alvarez, Y. Lecun, T. Gevers y A. M. Lopez, «Semantic road segmentation via multi-scale ensembles of learned features,» *Computer Vision-ECCV 2012. Workshops and Demonstrations*, pp. 586-595, 2012.
- [12] Y. Wang, E. K. Teoh y D. Shen, «Lane detection using B-snake,» *Information Intelligence and Systems, 1999*, pp. 438-443, 1999.
- [13] K.-M. Cheng, C.-Y. Lin, Y.-C. Chen y T.-F. SU, «Design of vehicle detection methods with OpenCL programming on multi-core systems,» *The 11th IEEE Symposium on Embedded Systems for Real-time Multimedia*, pp. 88-95, 2013.
- [14] K. Kluge y S. Lakshmanan, «A deformable-template approach to lane detection,» *Intelligent Vehicles '95 Symposium., Proceedings of the* , pp. 54-59, 1995.
- [15] Y. Wang y E. K. Teoh, «Lane detection using catmull-rom spline,» *IEEE International Conference on Intelligent Vehicles*, pp. 51-57, 1998.
- [16] A. Parajuli, M. Celenk y H. B. Riley, «Robust lane detection in shadows and low illumination conditions using local gradient features,» *Open Journal of Applied Sciences*, 3(01), n° 68, 2013.
- [17] T. Naito, T. Ito y Y. Kaneda, «The Obstacle Detection Method using Optical Flow Estimation at the Edge Image,» *2007 IEEE Intelligent Vehicles Symposium* , pp. 817-822, 2007.
- [18] T. N. Tan, «Efficient image gradient based vehicle localization,» *IEEE Transactions on Image Processing*, vol. 9, n° 8, pp. 1343-1356, 2000.
- [19] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer Science+Business Media, 2006.
- [20] H. Mori y N. Charkari, «Shadow and rhythm as sign patterns of obstacle detection,» *Industrial Electronics, 1993. Conference Proceedings, ISIE'93 - Budapest., IEEE International Symposium on* , pp. 271-277, 1993.

- [21] D. P. Y. Lee, «Robust vehicle detection based on shadow classification,» *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 3, pp. 1167-1170, 2006.
- [22] M. Bertozzi, A. Broggi y A. Fascioli, «Stereo inverse perspective mapping: theory and applications,» *Image and vision computing*, vol. 8, n° 16, pp. 585-590, 1998.
- [23] A. M. Muad, A. Hussain y S. A. Samad, «Implementation of inverse perspective mapping algorithm for the development of an automatic lane tracking system,» *TENCON 2004. 2004 IEEE Region 10 Conference*, vol. A, pp. 207-210, 2004.
- [24] S. Kay y G. Lemay, «Edge detection using the linear model,» vol. 9, pp. 577-580, 1984.
- [25] R. C. Gonzalez y R. E. Woods, *Digital Image Processing (3rd Edition)*, Prentice-Hall, 2008.
- [26] N. Otsu, «A Threshold Selection Method from Gray-Level Histograms,» *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, n° 1, pp. 62-66, 1979.
- [27] A. P. Dempster, N. M. Laird y D. B. Rubin, «Maximum Likelihood from Incomplete Data via the EM Algorithm,» *Journal of the Royal Statistical Society, Series B*, vol. 1, n° 39, pp. 49-58, 1977.
- [28] E. Maggio y A. Cavallaro, «5.2.1.1 Kanade-Lucas-Tomasi (KLT) tracker,» de *Video Tracking*, Singapore, Wiley, 2011, pp. 90-93.
- [29] H. C y S. M, «A combined corner and edge detector,» *Proceedings of the 4th Alvey Vision Conference*, pp. 147-151, 1988.
- [30] L. Bretzner y T. Lindeberg, «Feature Tracking with Automatic Selection of Spatial Scales,» *Computer Vision and Image Understanding*, n° 71, pp. 385-392, 1998.
- [31] Hough y P. V. C, «Method and means for recognizing complex patterns». United States of America Patente US3069654 A, 18 Dic 1962.
- [32] J. Canny, «A Computational Approach to Edge Detection,» *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vols. %1 de %2PAMI-8, n° 6, pp. 679-698, 1986.
- [33] J. Berger, «Statistical decision theory and Bayesian analysis,» Springer-Verlag, New York, 1985.

Anexos

Anexo 1. Estadística Bayesiana y Funciones de verosimilitud

En el mundo de la estadística, que busca proporcionar una metodología para analizar adecuadamente la información que tenemos y decidir cómo actuar con los resultados obtenidos sobre un espacio completo, es importante entender el siguiente esquema de la figura Anexo 2.1, basado en uno similar en [33]:

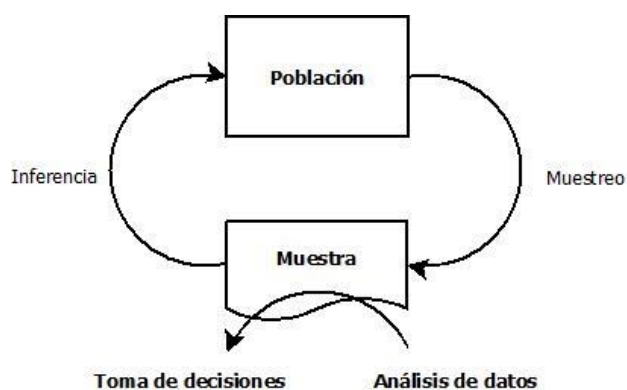


Figura Anexo 2.1 Esquema del funcionamiento de la estadística

La población es el conjunto de elementos sobre el que se realizan observaciones, mientras que la muestra será un número de elementos representativos de la población sobre los que se estudiará un determinado comportamiento.

Los resultados obtenidos del estudio de la muestra serán analizados y generalizados para toda la población.

La parte que afecta a este apartado es la de inferencia. La inferencia estadística es qué decisión tomamos con la información obtenida en la

muestra para aplicarla a la población. Actualmente, hay dos grandes tipos de inferencia [33]:

- La inferencia clásica, que supone que todos los sucesos son igual de verosímiles o posibles. La forma de calcularlo es dividir el número de casos probables en un suceso entre el número de casos totales.

- La inferencia subjetiva, que añade un juicio personal sobre la verosimilitud de que suceda un evento. En este tipo de inferencia se encuentra la estadística bayesiana, sustentado en el Teorema de Bayes.

De acuerdo a [33], la inferencia subjetiva cambia el concepto de probabilidad como frecuencia relativa de un suceso, porque dicho concepto puede variar, sobre todo a largo plazo. El nuevo enfoque será la probabilidad como el grado de convicción personal con el que un suceso puede ocurrir.

La ecuación más elemental que permite definir el grado de convicción de que un suceso tenga lugar de forma matemática dentro de la inferencia subjetiva es el teorema de Bayes, cuya expresión es:

$$P(\theta|S_i) = \frac{P(S_i|\theta)}{P(S_i)} P(\theta) \quad (2)$$

donde $P(S_i|\theta)$ es la probabilidad de que se dé una hipótesis S_i sabiendo una información θ .

La lectura principal de esta ecuación es que se transforma la hipótesis $P(\theta)$ definida mediante la probabilidad clásica y llamada *a priori* en este tipo de estadística, en la probabilidad $P(\theta|S_i)$ como probabilidad de inferencia subjetiva, llamada *a posterior*.

Al depender $P(S_i|\theta)$ del parámetro θ , variando la probabilidad de ocurrencia del suceso S_i en función de la elección de dicho parámetro, estaremos utilizando funciones de verosimilitud.

Anexo 1.1 Funciones de verosimilitud

En una función de verosimilitud, la probabilidad para un suceso dado depende de un parámetro, que por lo general será otro suceso. De otra manera, sea:

$$\mathcal{L}(\theta) = \mathcal{L}(S_i|\theta) = P_\theta(S_i)$$

donde $\mathcal{L}(\theta)$ es función del parámetro θ , S_i son las muestras posibles de un suceso y P_θ es la probabilidad para esas muestras cuando el parámetro es θ .

Por último, si clasificamos un suceso S_i de acuerdo al parámetro θ que maximiza la probabilidad $\mathcal{L}(\theta)$, estaremos utilizando el método de máxima verosimilitud. El parámetro θ que logra esto queda fijado como solución a la clasificación.

Anexo 1.2 Decisión bayesiana

Sean dos sucesos A y B dependientes entre sí, como la intensidad de un pixel A en función de su clase B . La probabilidad conjunta de ambos, más conocida como su intersección, puede desarrollarse como:

$$P(A \cap B) = P(A) * P(B|A)$$

De igual manera,

$$P(B \cap A) = P(B) * P(A|B)$$

Por lo que, al expresar ambas el mismo caso, que es la probabilidad de la intersección de ambos casos:

$$\begin{aligned} P(A \cap B) &= P(B \cap A) \stackrel{1}{\Rightarrow} \\ \stackrel{1}{\Rightarrow} P(A) * P(B|A) &= P(B) * P(A|B) \stackrel{2}{\Rightarrow} \\ \stackrel{2}{\Rightarrow} P(B|A) &= \frac{P(B) * P(A|B)}{P(A)} \end{aligned}$$

De manera que se demuestra la procedencia de la ecuación (2). Además, sabemos que, para el caso discreto:

$$P(A) = \sum_{i=0}^N P(A|B_i) * P(B_i) \text{ }^9$$

⁹ La forma más clara de ver esto sería con un diagrama de árbol.

donde N es el número de sucesos posibles como B . En nuestro ejemplo, sería el número de clases que hay, que serían 3: objeto, pavimento y línea.

Esto nos permite ver que el denominador normalizará la distribución. Como conclusión, la ecuación quedaría de la siguiente manera:

$$P(B|A) = \frac{P(B) * P(A|B)}{\sum_{i=0}^N P(A|B_i) * P(B_i)} \quad (3)$$

Si sustituimos $A = S_i$ y $B = \theta$, obtenemos:

$$P(\theta|S_i) = \frac{P(S_i) * P(S_i|\theta)}{\sum_{i=0}^N P(S_i|\theta) * P(\theta)}$$

Que se ajusta a la notación empleada en este anexo. Puede compararse con la ecuación (2).

Clasificar de acuerdo al parámetro S_i que maximiza la probabilidad $P(\theta|S_i)$ se conoce como máxima a posteriori.

Anexo 2. Kanade-Lucas-Tomasi (KLT)

El algoritmo KLT [28] permite determinar la posición de puntos de la imagen en instantes consecutivos, determinando entonces el desplazamiento que han experimentado, es decir, su movimiento, basándonos en el gradiente. Utiliza una zona de interés o *template* tomada de la imagen $k-1$, que nos servirá como un modelo de apariencia del entorno del píxel que está centrado a buscar en la imagen posterior, k .

Sea el área de búsqueda una región cuadrada con un tamaño fijo definido como $N = (2W - 1) \times (2W - 1)$. El problema de seguimiento (*tracking*) se reduciría a estimar la translación por movimiento de dicha región.

Para ello, definiendo:

$$x_k = \tilde{x}_k^{(0)} + \Delta x_k$$

donde x_k es el desplazamiento en el instante k para un punto dado, $\tilde{x}_k^{(0)}$ el candidato inicial como desplazamiento para el instante anterior $k - 1$ y Δx_k como un pequeño desplazamiento añadido producto del salto de una imagen a otra en un instante muy pequeño.

Con ello, podemos decir que:

$$I_k(w) = I_t(w - x_k) + n_k(w) = I_t(w - \tilde{x}_k^{(0)} + \Delta x_k) + n_k(w)$$

en el que $I_k(w)$ es el eje de coordenadas de la imagen actual y ventana de tamaño w con $|w - x_k| < W$, e $I_t(w - x_k)$ es el eje de coordenadas del *template* para la ventana w desplazada x_k .

De acuerdo a la ecuación, el desplazamiento Δx_k que minimice el error entre ejes coordenados será el que tomaremos como resultado para el algoritmo. Así, tomando $\epsilon(\Delta x_k)$ como el mínimo error cuadrático medio dependiente de Δx_k por lo dicho anteriormente:

$$\epsilon(\Delta x_k) = \sum_{|w-x_k|<W} \left[I_t(w - \tilde{x}_k^{(0)} + \Delta x_k) - I_k(w) \right]^2$$

Para calcular esto de una manera rápida, y dado a que los valores de Δx_k van a ser muy pequeños, el algoritmo se puede desarrollar por Taylor:

$$\epsilon(\Delta x_k) = \sum_{|w-x_k|<W} \left[I_t(w - \tilde{x}_k^{(0)}) - I_k(w) - b(w)' \Delta x_k \right]^2$$

Por último, resolviendo la función cuadrática con $\frac{\delta \epsilon(\Delta x_k)}{\delta \Delta x_k}$, tenemos:

$$\Delta x_k = \frac{\sum_{|w-x_k|<W} \left[I_t(w - \tilde{x}_k^{(0)}) - I_k(w) \right] b(w)}{\sum_{|w-x_k|<W} b(w)' b(w)}$$

Esta ecuación puede no converger, por lo que será necesario sustituir $\tilde{x}_k^{(0)}$ hasta converger.